Testing single-grain quartz OSL methods using sediment samples with independent age control from the Bordes-Fitte rockshelter (Roches d'Abilly site, Central France)

Thomsen, Kristina Jørkov; Murray, Andrew Sean; Buylaert, Jan-Pieter; Jain, Mayank; Helt-Hansen, Jakob; Aubry, T.

Published in:
Quaternary Geochronology

Link to article, DOI:
10.1016/j.quageo.2015.11.002

Publication date:
2016

Document Version
Peer reviewed version

Elsevier Editorial System(tm) for Quaternary Geochronology

Manuscript Draft

Manuscript Number: QUAGEO-D-13-00083R4

Title: Testing single-grain quartz OSL methods using sediment samples with independent age control from the Bordes-Fitte rockshelter (Roches d'Abilly site, Central France)

Article Type: Research Paper

Keywords: OSL; dating; single-grain; quartz; 14C age control

Corresponding Author: Dr. Kristina Jørkov Thomsen, PhD

Corresponding Author's Institution: DTU

First Author: Kristina Jørkov Thomsen, PhD

Order of Authors: Kristina Jørkov Thomsen, PhD; Andrew S Murray, PhD; Jan-Pieter Buylaert, PhD; Mayank Jain, PhD; Jakob H Hansen, PhD; Thierry Aubry, PhD

Abstract: We present quartz single-grain dose distributions for four well-bleached and unmixed sediment samples with independent age control (22-48 ka), from the archaeologically important Bordes-Fitte rockshelter at Roches d'Abilly, France. This site has previously been dated using 14C AMS dating and standard multi-grain OSL dating using both quartz and feldspar. The effect of rejection criteria usually employed in single-grain dating on dose and over-dispersion is tested using both laboratory irradiated samples and natural samples. It is shown that had these samples been analysed in the absence of other age control, standard modelling decisions based on the shape of single-grain dose distributions would have led to significant misinterpretation of results and a corresponding >40% underestimation in age. If we instead ignore this standard decision process and apply weighted average and mixing models then the most likely results deviate from the expected ages by >10%. Finally, we show that by careful consideration of the luminescence characteristics of individual grains, we are able to obtain good agreement with the independent age control by applying alternative rejection criteria but this is at the cost of reducing the accepted grain population by more than an order of magnitude, with the corresponding inevitable decrease in precision.

Testing single-grain quartz OSL methods using sediment samples with independent age control from the Bordes-Fitte rockshelter (Roches d'Abilly site, Central France).

K.J. Thomsen[a,*], A.S. Murray[b], J.P. Buylaert[a,b], M. Jain[a], J.H. Hansen[a], T. Aubry[c]

[a] *Center for Nuclear Technologies, Technical University of Denmark, DTU Risø Campus, DK-4000 Roskilde, Denmark*
[b] *Nordic Laboratory for Luminescence Dating, Department of Geoscience, Aarhus University, DTU Nutech, Risø Campus, DK-4000 Roskilde, Denmark*
[c] *Côa Parque, Fundação para a Salvaguarda e Valorização do Vale do Côa. Rua do Museu. 5150-610 Vila Nova de Foz Côa. Portugal*

* Corresponding author: krth@dtu.dk

Abstract

We present quartz single-grain dose distributions for four well-bleached and unmixed sediment samples with independent age control (22-48 ka), from the archaeologically important Bordes-Fitte rockshelter at Roches d'Abilly, France. This site has previously been dated using ^{14}C AMS dating and standard multi-grain OSL dating using both quartz and feldspar. The effect of rejection criteria usually employed in single-grain dating on dose and over-dispersion is tested using both laboratory irradiated samples and natural samples. It is shown that had these samples been analysed in the absence of other age control, standard modelling decisions based on the shape of single-grain dose distributions would have led to significant misinterpretation of results and a corresponding >40% underestimation in age. If we instead ignore this standard decision process and apply weighted average and mixing models then the most likely results deviate from the expected ages by >10%. Finally, we show that by careful consideration of the luminescence characteristics of individual grains, we are able to obtain good agreement with the independent age control by applying alternative rejection criteria but this is at the cost of reducing the accepted grain population by more than an order of magnitude, with the corresponding inevitable decrease in precision.

1 Introduction

Over the past 20 years standard optically stimulated luminescence (OSL) quartz dating using multi-grain aliquots has proved to be a powerful technique to establish absolute chronologies for sediments deposited during the past ~100 ka (see e.g. the review by Murray and Olley, 2002) for a range of depositional environments (aeolian, fluvial, colluvial etc). The fundamental assumption in this technique is that the latent OSL signal was completely reset to zero ('well-bleached') at the time of deposition. If this assumption is not met, standard multi-grain OSL dating is likely to result in an over-estimation of the true burial age because any residual signal will contribute to the measured dose. One approach to determining whether a sample is likely to have been well-bleached at burial is to measure so-called modern analogues from the same

depositional environment (these are recently deposited samples likely to have the same OSL and bleaching characteristics as the fossil sample of interest). Murray et al. (2012) reviewed the average OSL multi-grain dose estimates from 67 quartz modern analogues extracted from a variety of fluvial and colluvial environments and found an average dose off-set of ~2 Gy, indicating that in such environments incomplete bleaching is only likely to be of significance for younger samples (e.g. <20 ka). In such incompletely bleached samples, single-grain OSL dating may be preferred to maximise the probability of measuring aliquots consisting solely of well-bleached grains (e.g. Olley et al., 1999). The main challenge for such samples is then to identify the grains that are most likely to have been well-bleached at deposition. This is usually achieved using statistical models (e.g. Galbraith et al., 1999; Thomsen et al., 2007; Pietch, 2009) whose outcome is highly dependent on the uncertainties assigned to individual dose estimates. In general, single-grain quartz OSL dating has been shown to provide both accurate and precise burial ages for samples less than 20 ka (e.g. Feathers, 2003; Olley et al., 2004a; Arnold et al., 2009). However, single-grain quartz OSL dating is also extensively used for samples older than 20 ka (e.g. Jacobs et al., 2008a; Armitage et al., 2011; Arnold and Roberts, 2011; Arnold et al., 2013) where incomplete bleaching is expected to be of lesser importance despite the fact that only very few samples with independent age control have been reported in this age range. Olley et al. (2004b) dated seven deep sea sediments using single grains of quartz moved by wind 60 km from Australia across the Indian ocean and found that at least one sample (OSL6) required minimum age modelling to agree with independent age control (^{14}C giving an age of 28.9-32.6 ka). This was particularly surprising because their result suggested a very large (for an aeolian sediment) residual signal at deposition (equivalent to ~10 Gy). Demuro et al. (2012) found relatively good agreement between single-grain weighted average ages for two glaciofluvial gravels bracketing Reid till at Ash Bend giving ages of ~125 ka; they did not discuss whether equally good correspondence could be obtained using multi-grain dating. In general, the rationale for using single-grain OSL measurements for samples older than 20 ka is that grains with aberrant OSL characteristics (making them unsuitable for dating) can be discarded (e.g. Roberts et al., 1999; Yoshida et al., 2000; Jacobs et al., 2003; 2011; Arnold and Roberts, 2011) and that sediment mixing, beta dose rate heterogeneity, roof spall contamination or other sources of inhomogeneity (Jacobs, 2010, Jacobs et al., 2012) can be identified and any contaminating grains excluded prior to age estimation (e.g. Roberts et al., 1998; Jacobs et al., 2003; Feathers et al., 2006; Porat et al., 2006). Jacobs et al. (2011) and others have argued that both the precision and accuracy of OSL ages can be improved by using single-grain measurements.

The main purpose of this study is to investigate the accuracy of single-grain OSL dating beyond 20 ka using four samples with associated ^{14}C ages from the Bordes-Fitte rockshelter (Roches d'Abilly site, Central France) at which a succession of Middle and Early Upper Palaeolithic occupations is preserved. This site was originally dated (Aubry et al., 2012) using standard multi-grain quartz OSL measurements, the $pIRIR_{290}$

signal from K-rich feldspar (KF) and ^{14}C AMS on bone and tooth collagen using an ultrafiltration protocol (Lyon and Oxford Laboratories) or extraction with alkali (Beta Analytic). These results were later revised and expanded in Aubry et al. (2014). The different dating methods gave stratigraphically consistent results in the range 22-48 ka, with the only exceptions being a few of the ^{14}C ages obtained by extraction with alkali (Aubry et al., 2012). This was not considered surprising, because an alkali-extraction ^{14}C AMS age obtained on bone or tooth sample beyond 35 ka should be interpreted as a minimum age (Jacobi et al., 2006; Higham et al., 2006). The agreement of the multi-grain aliquot quartz and KF ages with each other and the consistency with the associated ^{14}C ages obtained by ultrafiltration gives us confidence in the accuracy of the overall ages of the occupation layers, and also in the accuracy of the multi-grain luminescence ages. As a result these samples were selected for single-grain testing.

In the present paper we describe the quartz multi- and single-grain luminescence characteristics of these samples and present the measured natural single-grain dose distributions. Standard multi-grain ages are derived and used to support the age control. Published approaches to selection/rejection and statistical analysis are applied to the single-grain data, and the results compared to multi-grain quartz, feldspar and ^{14}C ages. It is shown that these single-grain dose estimates underestimate the expected values. Finally, we argue for the use of additional or alternative rejection criteria to produce dose estimates unbiased by single-grain OSL signals either close to or in saturation, or with slow decay rates.

2 Samples and contexts

The four OSL samples were taken from the Bordes-Fitte rockshelter at Roches d'Abilly. The site is located in Central France on the southern margin of the Paris Basin, along the valley of the Creuse river, a tributary of the Vienne river which joins the Loire Valley on its south bank, downstream of Tours. The rockshelter is situated in a cliff on the south bank of the Creuse valley, just after the confluence with the Claise. In this region decalcification of the Upper Turonian formations and secondary silicification processes have provided abundant large nodules of the high-quality flint used to make stone-tools throughout prehistory. Middle and Upper Palaeolithic occupation is identified in caves, rockshelters, and open-air sites and by the Grand-Pressigny Neolithic large blade workshops, exploiting the specificity of local flint (Marquet, 2011).
The Roche d'Abilly site is a complex of *loci*, around a quarried escarpment with a 300-meter-long exposure of Upper Cretaceous carbonates (chalk, sandy marl, bio-calcirudite and calcarenite with chert bands and compact silicified limestone). These lithologies have been altered by weathering processes, resulting in the development of horizontal inter-bedded caves and rockshelters, subsequently filled by siliciclastic and carbonate-rich sediments. These provide protected environments where archaeological sediments and features are preserved. A collapsed rockshelter at the western end of the quarry was excavated in 1949, and the results summarily published by Bordes and Fitte (1950). Latter, André Chollet and Jean-Claude Marquet

carried out supplementary surveys, but both were called to a halt by the presence of large thick slabs of calcarenite, resulting from prehistoric roof collapse. In 2007, new archaeological investigations involved removing the large collapsed slabs of compact calcarenite and revealing the underlying intact Palaeolithic deposits; this was undertaken both near the Bordes-Fitte excavation's eastern limit, and also at other locations along the cliff (Aubry et al., 2012).

The lithostratigraphic sequences observed are composed of near-surface sedimentary facies with vertical and lateral variations, in a context dominated by run-off and gravitational sedimentary processes (Aubry et al., 2012). Field description and micromorphological analysis permit the reconstruction of several episodes of sediment slope-wash and endokarst dynamics, with hiatuses and erosional phases. The Bordes-Fitte rockshelter archaeostratigraphic sequence includes Châtelperronian artefacts, inter-stratified between Middle Palaeolithic and Aurignacian occupations, and represents an important contribution to the debate about the characterization and timing of the changes in material culture associated with the Middle-to-Upper Palaeolithic transition between 50 and 40 ka in Western Europe and their relationship with the Neanderthal and Anatomically Modern Human populations and behaviours (Aubry et al., 2012).

Refitting has been systematically applied to lithic assemblages from the site independent of their stratigraphic provenience. Refits are essentially from remains recovered from the same stratigraphic unit, but with some exceptions (see below).

The samples were collected over an area of approximately 2.5 by 2.5 metres and are stratigraphically connected in three dimensions (see Figure 1 and Figure S1). Published ^{14}C ages for this site (Aubry et al., 2012; 2014) are summarised in Table 1. These have been recalibrated where necessary using the OxCal v4.2 calibration program (Bronk Ramsey, 2013). Samples 092201 and -02 were collected from unit GFU D2 and are associated with five ultrafiltration ^{14}C ages, in the ~46 to 40 ka interval (OxA-22316, -22342, -226470, -226471, -226472). This stratigraphy is supported by a further ^{14}C age (OxA-29527) of ~46 ka taken from the underlying unit GFU D1 at a distance of ~2.5 m from section A (see Figures 1 and S1). Another age (Beta-249595) of ~36 ka was obtained for a bone diaphysis recovered in GFU D2, using collagen extracted with alkali preparation. Stratigraphically this sample should clearly be older; the most probable explanation for this inconsistency is incomplete decontamination of recent carbon during the Langin pre-treatment (Langin, 1971); it is well-known that the ages of most samples older than 35 ka processed using this conventional pretreatment are underestimated (Higham, 2011). This ^{14}C sample (Beta-249595) is not considered further. Within unit GFU D2, there is no stratigraphic order in the remaining five ^{14}C ages (OxA-22316, -22342, -226470, -226471, -226472) and they are indistinguishable before calibration (See Table 1). The arithmetic average uncalibrated age is 39.9±0.7 ka (n=5) and the weighted age is 39.6±0.7 (CAM_{UL}, Arnold et al., 2009), with an over-dispersion of 3±2%. The resulting calibrated ages are 44964-42610 yr and 44789-42435 yr, respectively. We take 44964-42610 yr as the best independent estimate of the age of unit GFU D2.

The two OSL samples (092201 and -02) were taken within 10 cm of each other laterally and within 5 cm vertically. The two ^{14}C samples (OxA-22316 and OxA-22342) were taken between 50 and 100 cm away at the same stratigraphic level. Thus, we expect these OSL samples to have been deposited ~44 ka ago.

At least two human occupations (Discoidal and Châtelperronian) characterized by distinct lithic flake and blade production could be distinguished in GFU D2. No contamination of GFU D2 by Aurignacian material from overlying layers was detected, either as individual artefacts or as cross-unit refits. However, where unit GFU D2 was not capped by GFU D3 or rock fragments, some Châtelperronian artefacts have been incorporated in the bottom of the overlying GFU E during its deposition in channel-fill along the slope and around the collapsed roof fragments. Alteration variables reveal that Middle Palaeolithic and Châtelperronian remains of unit GFU D2 were not reworked at a large scale, nor were they exposed for a significant time after deposition (Aubry et al., 2014). Unit GFU D is largely capped by rockfall from the shelter roof in the northern part of the excavated area (Figure 1), although roof fragments from this first collapse did not directly cap our two samples 092201 and -02. However, both these OSL samples and two ^{14}C samples (OxA-22316 and OxA-22342) are capped by GFU D3 (section B in Figure 1). The three ^{14}C samples taken from GFU D2 in section A of Figure 1 are directly capped by rockfall. As a result all these samples are regarded as stratigraphically secure.

Unit GFU E yielded ^{14}C ages in the range 39 to 40 ka (Lyon-6920 and Beta 249596). In those parts of unit E not covered by roof fragments, the presence of some Châtelperronian artefacts and cross unit refits with unit GFU D2, taken together with a pattern of orientation similar to that of the Aurignacian lithic artefacts in unit GFU E suggests some erosion and re-deposition of unit GFU D2 before the deposition of unit GFU E. However, no Châtelperronian artefacts or refits have been observed in the immediately overlying unit GFU F from which OSL samples 092203 and -04 were taken. Aurignacian remains from unit GFU E are sometimes present at the base of the GFU F unit, although the sharpness of the stratigraphic boundary between these units suggest that such upwards mobility is less likely for sediment grains. A ^{14}C age of ~36 ka (Beta-234193) was obtained for a bone diaphysis sample collected at the interface of GFU E and F; this age may be an underestimate (in a similar manner to beta 249595), but it cannot be older than the age of unit E, which is based on both an ultra-filtration and collagen ^{14}C age and so there is very little room for underestimation in the age of the E/F interface. A red deer antler sample collected within the overlying GFU F deposit gave an age of ~23 ka (OxA-22315). This ^{14}C sample was taken at the same depth and within 10 cm of OSL sample 092203 and in our view provides a reliable age control for sample 092203. Sample 092204 was taken from the same unit GFU F but at a lateral distance of approximately 2.5 m in the sediment overlying the Beta-234193 ^{14}C sample. Unit GFU F is in turn capped by rocks from a second shelter roof collapse event. Thus, we would expect that these OSL samples (092203 and -04) must be significantly younger than ~36 ka and that 092203 should be consistent with the ^{14}C age of sample OxA-22315 (~23 ka).

Sample 092204 might be of the same age or older. The details of these ^{14}C ages are given in Aubry et al. (2012, 2014).

To summarise these stratigraphic relationships, samples 092201 and -02 are taken from unit GFU D2 and should be of similar age, between 45.0 and 42.6 ka. They should also be older than the overlying unit GFU E (37-41 ka). Samples 092203 and -04 from unit GFU F should be younger than unit GFU E. Sample 092203 should be ~23 ka and -04 maybe similar or older, but younger than the unit GFU E/F boundary at ~35-36 ka.

Sample 092201 was taken ~15 cm above bedrock and sample 092202 ~20 cm above. Sample 092203 lies ~15 cm below a capping rock and ~15 cm above the corner of a prior rock fall. Sample 092204 lies only ~5 cm below a capping rock and 15 cm above bedrock. Because of this we also collected four angular rock fragments (not directly associated with any of the OSL samples) to provide an estimate of the gamma dose rates presumed to derive from the over- and underlying rocks.

3 Experimental details

Purified quartz and K-rich feldspar grains (180-250 μm) were extracted from the sediment by sieving, acid cleaning using HCl, H_2O_2, heavy liquid separation (2.58 g cm^{-3}) and finally etching of quartz in 40% HF and feldspar in 10% HF (Aitken, 1985; 1998). The quartz purity was tested using IR stimulation to detect the presence of feldspar and no significant IR signals were observed. Nevertheless, the IR depletion ratios (Duller, 2003) of multi-grain aliquots are further discussed in the supplementary information (see section S3).

3.1 Instrumentation

All multi-grain OSL signals were measured using automated TL/OSL Risø DA-20 readers (Bøtter-Jensen et al., 2010) equipped with arrays of blue (470±30 nm) and infrared (IR, 870±40 nm) stimulation LEDs providing stimulation powers at the sample position of approximately 80 and 130 mW/cm^2, respectively. EMI 9635QA photomultipliers in combination with 7.5 mm Hoya U-340 filters were used to detect the quartz OSL signals. All single-grain OSL measurements were done using Risø single grain laser attachments (Bøtter-Jensen et al., 2003); these use a 10 mW Nd:YVO_4 solid-state diode-pumped laser emitting at 532 nm as the stimulation light source. The laser is focused (~20 μm spot size) sequentially onto a 10 by 10 array of grain holes (depth and diameter of 300 μm) spaced on a 600 μm square grid on 9.7 mm diameter aluminium sample discs. All single-grain discs were screened for contamination prior to measurement.

In situ beta irradiations used calibrated $^{90}Sr/^{90}Y$ beta. To ensure that the scatter observed in the single-grain dose distributions does not result from beta source inhomogeneity (e.g. Spooner and Allsop,

2000) the approach developed by Lapp et al. (2012) to correct for spatial non-uniformity was applied. However, the beta sources used for single-grain irradiations all had good spatial uniformity (<5% standard deviation across the sample area) and thus correcting beta source non-homogeneity did not result in significant changes to dose or scatter.

Gamma irradiations employed a ^{60}Co gamma cell. The samples were placed in glass tubes with an internal diameter of 5 mm and a wall thickness of 2 mm. The dose rate of 7.67±0.01Gy min^{-1} delivered by the gamma cell was calibrated using EPR measurements of Alanine (see supplementary information S2 for details). The quartz samples were given doses of 4.9, 9.5, 28.1, 74.8 and 107.4 Gy, respectively. It is expected that the dose rate varied by less than 3% across the sample volume (see section S2).

Dose rates were determined using high-resolution gamma spectrometry calibrated as describe in Murray et al. (1987) and the conversion factors of Guérin et al. (2011). Crushed and homogenised samples were mixed with wax and cast in a fixed cup-shaped geometry. To ensure equilibrium between ^{222}Rn and ^{226}Ra, these casts were then stored for a minimum of 20 days (>5 ^{222}Rn half-lives) before counting. We assume that the present day burial depths represent the life time burial depths and cosmic ray contributions are based on Prescott and Hutton (1994). Life time water contents are based on measured (field) values.

3.2 OSL measurements

The single-aliquot regenerative-dose (SAR) procedure (Murray and Wintle, 2000) was used for equivalent dose determination. A double SAR procedure (Banerjee et al., 2001) employing IR stimulation prior to blue or green stimulation (at 125 °C) was used on all samples to minimize the effects of any feldspar contamination. For both multi-grain and single-grain measurements the prior IR stimulation was at 125 °C for 40 s (IR stimulation at this temperature will result in a depletion of the fast component of ~0.2%, Jain et al., 2005). For multi-grain measurements blue stimulation at 125°C for 40 s was used whereas for single-grain measurements green laser stimulation was at 125 °C for 0.9 s. A preheat temperature of 260 °C for 10 s and a cutheat temperature of 220 °C was used for all dose measurements unless otherwise specified (e.g. preheat plateau measurements in section 6.2). A high-temperature blue bleach at 280 °C for 40 s was inserted in between each SAR cycle to minimise potential recuperation effects (Murray and Wintle, 2003). For multi-grain measurements the signal was summed over the initial 0.5 s of stimulation and the background was summed over the subsequent 0.5 s of stimulation (early background subtraction, EBG). For single-grain measurements the signal was summed over the initial 60 ms and the background was summed over the final 150 ms (late background subtraction, LBG). Multi-grain aliquots were prepared in stainless steel cups using an ~5 mm spot of silicon oil. Up to 100 individual grains were loaded into each of the single grain discs referred to above. Visual inspection under subdued red light using a microscope ensured that a maximum of one grain was placed in each single grain hole.

Multi-grain laboratory dose response curves (DRCs) consist of a minimum of three sensitivity corrected regeneration dose points as well as a recuperation point (i.e. zero dose point) and a recycling point. Single-grain response curves used a minimum of 6 regeneration points. All DRCs have been fitted using a single saturating exponential passing through the origin, i.e. $L_x/T_x = a \times [1-\exp(-D/D_0])$, where L_x/T_x is the sensitivity corrected OSL response, a is the saturation value and D_0 is a measure of the curvature of the DRC. Doses are derived by interpolating the natural sensitivity correct signal onto the DRC and uncertainties are assigned to these individual doses using "Analyst 4.00" (Duller 2007) and consist as a minimum of contributions from Poisson counting statistics and curve fitting errors.

3.3 OSL rejection criteria

In the first part of the data analysis we apply standard rejection criteria to individual aliquots: (i) the relative uncertainty on the first (natural) test dose signal must be less than 20% ("s_{TN}<20%"; note that this rejection criterion is stricter than the commonly employed T_N> 3 times the standard deviation of the background), (ii) the recycling ratios must be consistent with unity within 2 standard deviations ("Recyc."), (iii) the IR depletion ratio (Duller 2003) with sensitivity correction must be consistent with unity within 2 standard deviations ("IR depl.") and (iv) the recuperation dose must be smaller than or consistent with 1 Gy ("Recup."). We choose to convert the recuperation sensitivity corrected signal into a dose by interpolation onto the aliquot specific DRC forced through the origin (see above) and use an absolute threshold value of 1 Gy (as opposed to a relative threshold value of typically 5%) in an attempt to avoid biasing the measured dose distribution by preferentially removing low dose estimates. In addition to these rejection criteria we also exclude all aliquots with natural sensitivity corrected signals (L_n/T_n) appearing to be in or above saturation of the laboratory measured DRCs, i.e. $L_n/T_n+s_n>a$, where s_n is the uncertainty (based on Poisson statistics) assigned to L_n/T_n. Application of this latter criterion clearly has the potential to bias the resulting dose distributions, and this problem is considered further in section 8.4. For the multi-grain aliquot natural measurements of the four samples a total of 3 aliquots (out of the 256 measured) were in saturation (~1.2%). For single-grain measurements of the two older samples (092201 and -02) ~22% of the otherwise accepted aliquots were rejected due to saturation. The corresponding number for the two younger samples (092203 and -04) is ~11% (see Table 4).

However, application of these standard quartz OSL rejection criteria does not significantly change either the mean dose or the dispersion (see section S3 for further details on the multi-grain data). In Figure 2 we show the effect of applying the rejection criteria to nine single grain dose distributions (both dose recovery and natural dose distributions). All results have been normalised to the result obtained for all grains passing the s_{TN} <20% criterion. The main effect of applying these rejection criteria is to reduce the number of accepted grains by approximately 30%. Neither the central dose (CAM, Galbraith et al., 1999) nor the relative over-dispersion (OD) is significantly changed by applying the standard rejection criteria. The same

observation was made by Thomsen et al. (2012). Despite the fact that application of the standard rejection criteria does not alter the individual dose distributions, we have chosen to use the dose distributions resulting from application of these rejection criteria to make this study directly comparable with other published studies unless otherwise stated. Nevertheless, we emphasise that application of these criteria does not result in a less dispersed data set.

4 Dose rate measurements

The radionuclide concentrations derived from the four OSL samples (092201-04) and the four rock fragments (TA2265R-68R) are summarised in Table 2. All sediment activity concentrations are relatively high compared to the corresponding values from the rock samples. Parent ^{238}U is not precisely measured in our gamma spectrometry facility; for dosimetry calculations we assume ^{234}U and ^{230}Th concentrations midway between the observed values of ^{238}U and ^{226}Ra with uncertainties equal to those of the ^{238}U analyses. For the sediment samples, we also assume a 20±10% loss of ^{222}Rn compared to its parent ^{226}Ra. The rock samples are more likely to be closed systems, and so for calculation of the gamma dose rate from surrounding rock clasts, we assume the U-series to be in secular equilibrium with a concentration given by ^{226}Ra. Table 2 lists the resulting infinite matrix dry beta and gamma dose rates.

The dose rates to all samples need modification to take into account heterogeneity in the gamma field. In the worst case (092204) the capping rock lies ~5 cm above the OSL sample, which in turn is ~15 cm above bedrock. This implies a contribution of ~29% of the rock gamma dose rate to the total dose rate (based on Aitken, 1985, Table H1; average gamma dose rate 0.20±0.02 Gy/ka based on Table 2). Since the infinite matrix rock gamma dose rate is only ~20% of the infinite matrix sediment gamma dose rate, the total gamma dose rate decreases by ~20%, and the rocks only contribute ~7% of the total gamma dose rate. For sample 092203 the contribution from the rock gamma dose rate is only ~2% of the total gamma dose rate, and this is ~10% lower than the infinite matrix sediment gamma dose rate. For samples 092201 and -02, the infinite matrix gamma dose rate is reduced by <5% due to the proximity of the less radioactive underlying bedrock.

The total dose rates given in Table 2 are based on these modelled gamma dose rates and include a cosmic ray contribution (calculated using the lifetime burial depths assumed equal to the sampling depths given in Table 2); realistic shielding corrections have a negligible effect on the total dose rate. Sediment water content corrections are also included, based on the presumed lifetime average water content shown in column 3. These lifetime averages are based on observed water contents at the time of sampling. The site is at the top of a steep slope and is well-drained; it is unlikely that the water contents were ever significantly higher than today for any prolonged period. Nevertheless, we have associated a ±4% uncertainty on these water contents in our dose rate calculations. At two standard deviations this allows for the possibility of lifetime average water contents >10% by weight and these are considered very unlikely.

Note that these dose rates are slightly different from those summarised Aubry et al. (2012; 2014), mainly because we have used the more recent conversion factors given by Guérin et al. (2011) and because we have revised the modelling of the gamma dose rates.

The possible effects of the uncertainties in some our dosimetry assumptions of the total dose rates are now considered. Because ^{238}U is poorly known, examination of the state of equilibrium of individual samples is not useful. However, the weighted average of the $^{226}Ra/^{238}U$ activity ratio in the sediments is 1.21±0.14 (n=4), and we conclude that we cannot confidently deduce the presence of any significant disequilibrium between ^{238}U and ^{226}Ra. Nevertheless, it is interesting to consider the effect of a ~20% disequilibrium on the total dose rates. Assuming that, for each sample, the first part of the uranium series down to and including ^{230}Th is given by the ^{238}U analysis, then the total dose rate would, on average, increase by <1%. Assuming that this disequilibrium step lies between ^{234}U and ^{230}Th would also increase the dose rate by <1%. These negligible deviations would apply if these disequilibria are supported through time. If, on the other hand, these possible disequilibria were unsupported (i.e. changing with time according to the relevant half-lives) and the disequilibrium was initially established between ^{234}U and ^{230}Th at the time of burial, then because these samples are all < 50 ka (see Table 1 and 3), ^{230}Th (and daughters) would have decayed by <40% since deposition; this would result in a <1.5% increase, on average, in total dose rate. Given the relatively short half-life of ^{226}Ra (~1600 years) compared to the site lifetime, it is not considered likely that the ^{226}Ra excess (if it exists) could have been unsupported throughout burial – this would imply a decay of over >30 lifetimes, and an initial ^{226}Ra excess of >5 $GBq.kg^{-1}$. It is also conceivable, although unlikely, that the possible 20% ^{226}Ra excess was acquired at time of excavation, so that during the life time of the site the first part of the U-series was in equilibrium. In that case the total dose rate would decrease by <2% on average.

There is considerable evidence in the literature that ^{222}Rn can escape from sediments at considerable depth (e.g. Tanner, 1980; Olley et al., 1997); this is why we have chosen an average ^{222}Rn to ^{226}Ra ratio of 0.8±0.1 (see above). Nevertheless, it is instructive to consider the effect of assuming that all ^{222}Rn is retained. In that event the average total dose rate would increase by 2.5%; this is small compared to other sources of uncertainty.

In summary, the likely effects of U-series disequilibrium are negligible at this site, and the water contents are low and likely to have been so throughout the site lifetime. Burial depth is relatively well defined (because of the roof of the shelter). We conclude that there are unlikely to be any major sources of unquantified uncertainty in these dose rates.

5 Predicted quartz doses

Aubry et al. (2012) extracted quartz and K-rich feldspar fractions and measured these using OSL from multi-grain aliquots from four samples (except that only quartz was available for sample 092201). Results are summarised in Table 3; the feldspar doses (measured using the $pIRIR_{290}$ signal, e.g. Thiel et al., 2011) are taken directly from Aubry et al. (2012), but additional quartz measurements have been made for our study. In Figure 3 we have converted the ^{14}C ages relevant to the OSL samples and multi-grain $pIRIR_{290}$ ages into predicted quartz doses using the quartz dose rates (see Table 2). For three OSL samples (092202, -03 and -04) we can directly compare multi-grain quartz and feldspar ages. Those from feldspar are systematically larger than those from quartz by, on average, 3±1 ka corresponding to an average dose off-set of 8±2 Gy. Given the two orders of magnitude difference in laboratory bleaching rates between the fast component from quartz and the $pIRIR_{290}$ signal from feldspar (Buylaert et al., 2012; Murray et al., 2012), and the almost certain presence of some difficult-to-bleach residual feldspar dose at deposition indicates that it is very likely that both minerals were effectively well-bleached at deposition. Since all the multi-grain quartz and feldspar doses are consistent with the ^{14}C intervals, we conclude that there is no evidence that our multi-grain quartz ages are inaccurate estimates of the last time these sediments were exposed to light. For simplicity throughout this paper we compare our various single-grain dose estimates with those derived from multi-grain quartz (given in Table 3), taking into account the uncertainties on the latter. From the discussion given above these doses are very unlikely to overestimate the true burial dose. However we recognise that it is possible that they may underestimate, although in our view this is unlikely, because of the agreement with the doses predicted from the feldspar ages. We reconsider the effects of this assumption when it has a direct effect on our later discussion (see section 8).

6 Quartz luminescence characteristics

6.1 Dose response and stimulation curves

Typical natural and regenerated normalised OSL stimulation curves measured using multi-grain quartz aliquots are compared with that from Risø calibration quartz (known to be fast component dominated) in the inset to Figure 4a. These stimulation curves are indistinguishable from each other indicating that the OSL signals from our samples are fast component dominated with no significant contribution from slower components. We routinely apply early background subtraction (EBG) to all multi-grain measurements, but for these samples a late background subtraction (LBG) could have been applied without significantly changing the dose estimate (e.g. the ratio between the D_e calculated using LBG and EBG for sample 092204 is 1.01). Figure 4b shows a frequency histogram of the D_0 values for all natural multi-grain aliquots for sample 092201. The average D_0 value is 108±3 (n=63) for this sample (the average D_0 values for samples

092202, -03 and -04 are 97±6, n=51; 95±3, n=75 and 93±3 Gy, n=67, respectively). Thus, based on the shapes of the dose response curves, we expect to be able to measure natural doses accurately up to approximately 200 Gy.

The inset to Figure 4d shows four natural stimulation curves from representative single grains from sample 092201. The grains have been selected to illustrate the differences in decay shape between individual grains; this is addressed further in section 8.4.3. The grey shaded bands indicate the LBG summation intervals used when estimating doses. In the literature, single-grain DRCs are consistently reported to vary considerably from grain to grain (e.g. Yoshida et al., 2000). Figure 4d shows examples of four representative single-grain normalised DRCs from sample 092201. The D_0 values are given next to each curve. Figure 4e shows a frequency histogram of all D_0 values known to better than 50%, for those grains accepted into the natural dose distribution for sample 092201. The D_0 values range between 17±10 and 1430±1930 Gy, where the large uncertainty on the latter reflects that the DRC for this grain appears to be effectively linear in the dose range under consideration (up to 250 Gy for this grain). The weighted average D_0 value is 102±4 Gy which is very similar to the arithmetic D_0 average of the multi-grain aliquots. The latter will tend to be dominated by the brighter grains, suggesting that the most sensitive grains have similar average D_0 to that of the less sensitive grains. Again we would expect that on average we should be able to measure natural doses up to ~200 Gy accurately.

6.2 Preheat plateau measurements

Figure 4c shows the dependence of the natural arithmetic average multi-grain dose on thermal pre-treatment (i.e. preheat temperature) for samples 092202 (squares) and 092203 (circles). In this experiment the thermal treatment (cutheat) prior to the test dose OSL was kept constant at 220 °C. Each point represents the average dose obtained from 8 aliquots. The measured natural doses do not appear to be sensitive to the choice of preheat temperature in the temperature interval considered here. In Figure 4f the result from four similar single-grain experiments using sample 092203 are shown. Each point represents the CAM average of more than 30 individual doses. For preheat temperatures less than 280 °C the measured dose is independent of the choice of preheat temperature. Based on this we have chosen to use a standard preheat temperature of 260 °C for 10 s and a cutheat temperature of 220 °C for all quartz OSL measurements.

6.3 Quartz dose recovery (standard rejection criteria)

Dose recovery experiments were undertaken to assess whether we are able to give a known dose in the laboratory prior to any thermal treatment and measure it accurately using our SAR measurement protocol. A dose recovery ratio within 10% of unity is generally considered to be satisfactory (Wintle and Murray, 2006). Single-grain dose recovery experiments were also undertaken to assess the variability in our dose distributions arising from intrinsic sources of uncertainty.

Dose recovery experiments were undertaken using both multi-grain and single-grain aliquots. Initial bleaching of the natural signal at ambient temperature was undertaken either using a daylight simulator (Hönle SOL 2) at a lamp/sample distance of 80 cm for 2 hr or using the blue LEDs in the TL/OSL reader twice for 100 s with an intervening pause of 10,000 s to allow charge optically transferred into the trap giving rise to the 110°C TL peak to decay. A known dose (either beta or gamma) ranging between 4.8 and 122 Gy was then given to each aliquot, and this was subsequently measured using our SAR protocol. The multi-grain and single-grain dose recovery results are summarised in Table S1 and Table 4, respectively. Dose recovery ratios (measured dose divided by given dose) are calculated using the arithmetic mean dose for multi-grain measurements and the CAM dose for single-grain measurements.

All dose recovery ratios are considered to be satisfactory, i.e. are indistinguishable from the range 0.9 to 1.1, for both multi-grain and single-grain measurements regardless of the method of bleaching (daylight simulator or blue LEDs). For example, for sample 092201 and blue bleaching we obtain dose recovery ratios of 0.95±0.02 (n=14) and 0.96±0.02 (n=165) for multi-grain and single-grain aliquots, respectively. For SOL2 bleaching the corresponding numbers are 0.92±0.04 (n=10) and 1.01±0.02 (n=103); it appears that we are unable to observe a significant systematic difference between the two methods of bleaching, in contrast to the observations by Choi et al. (2009).

Thus, based on both multi-grain and single-grain beta dose recovery experiments we conclude that we are able to measure accurately a laboratory dose ranging between 5 and 120 Gy, given before any prior thermal treatment.

7 Natural dose distributions

A total of 256 multi-grain aliquots were measured; 3 of these (~1%) were rejected because of saturation. The dose distributions of the two younger samples (092203 and -04) are approximately symmetrical but those of the two older samples may be slightly positively skewed (092201and -02, see Figure S4). The individual unweighted averages and CAM averages are consistent within 1 standard error (see Table 5). The relative ODs range between 17±2 and 26±2 % and, as expected, are completely consistent with the relative standard deviations, i.e. the contribution from counting statistics and curve fitting errors to the relative standard deviation is not detectable. For the multi-grain dose distributions it is clear that there is no advantage in deriving a CAM dose in preference to an average dose (see Table 5); in what follows we employ the average (arithmetic) multi-grain doses throughout.

Figure 5 shows the natural single-grain dose distributions. On average ~9% of the measured grains were accepted into each distribution. Uncertainties of individual dose estimates have been assigned based on counting statistics, curve fitting errors and an additional contribution of 15% to account for instrument reproducibility and other intrinsic sources of variability. We have chosen to assign this additional uncertainty

of 15% based on the smallest observed relative OD in our gamma dose recovery dose distributions (see Table 4 and section 8.4.1) when applying the standard rejection criteria given in section 3.3. Others (e.g. Arnold and Roberts, 2009; Arnold and Roberts, 2011) have also used an additional uncertainty of 15%, but in these cases because they argue that this is typical of the relative OD observed for natural well-bleached samples. (For comparison the dose distributions for samples 092201 and -03 are presented without the additional 15% uncertainty as radial plots in Figure S5.)

The natural dose distributions appear approximately symmetrical, although the CAM doses are approximately 7% smaller than the simple arithmetic averages. There is significant scatter in the individual dose distributions; these have relative ODs ranging between 31±2 and 38±3% with an average of 34±1% (this increases to 37±1% if the assigned uncertainties are instead based solely on counting statistics and curve fitting errors). For samples 092201 and -02, ~22% of the otherwise acceptable grains were rejected because the natural signal was in saturation. The corresponding number for samples 092203 and -04 is ~11%. Similar numbers of grains rejected due to saturation have been reported before in the literature (e.g. Jacobs et al., 2008a)

A number of different methods are usually applied to determine single-grain burial doses, e.g. CAM (Galbraith et al., 1999), CAM unlogged (CAM_{UL}; Arnold et al., 2009), various minimum age models (e.g. MAM, Galbraith et al., 1999) and the finite mixture model (FMM, Galbraith and Green, 1990). In the following we will apply these various statistical approaches and compare the results with the multi-grain quartz results; we know these to be accurate based on the ^{14}C age control and the agreement with feldspar ages (see Figure 3).

8 Single-grain burial doses

8.1 Minimum age modelling

It has often been argued (e.g. Olley et al., 2004b; Arnold and Roberts, 2009) that relative ODs of larger than 20% in single-grain dose distributions imply the presence of incomplete bleaching. (Note that this threshold value of 20% was suggested for single-grain dose distributions where the uncertainties assigned to individual dose estimates were based on counting statistics and instrument reproducibility alone.) The single-grain instrument reproducibility has been reported to result in an additional uncertainty of approximately 6% per dose estimate depending on the chosen signal summation interval (e.g. Thomsen et al., 2005). If we base our estimates of uncertainty on counting statistics, curve fitting errors and an instrument reproducibility of 6% per dose estimate then the average relative OD for our four samples is 36.3±1.3%, which is significantly larger than the typical 20% threshold – according to the published literature this would seem to indicate that our distributions are incompletely bleached, in which case minimum age models should be applied to identify the grains that were well-bleached at burial (Olley et al., 2004b). Bailey and Arnold (2006) have

since suggested a decision tree making use of the relative OD and the weighted kurtosis and skewness to assess which statistical model (CAM, lowest 5%, Olley et al., 1998; 1999, or MAM) should be applied to individual dose distributions to extract the burial doses. For our samples this decision tree predicts that the burial dose should be obtained using the lowest 5% (092202), MAM-3 (092203) and MAM-4 (092201 and 03). However, using the lowest 5% underestimates the multi-grain quartz dose by ~ 75%, MAM-3 by ~45% and MAM-4 by ~60%. The average underestimation obtained by applying MAM-3 to all four samples is ~42±2% (n=4; see Figure 3). Thus, it is clear that an OD larger than 20% does not necessarily imply that minimum age models should be applied and that the decision tree proposed by Bailey and Arnold (2006) does not result in accurate estimates of burial doses for these samples.

We have also tested this decision tree on the 28 single-grain dose recovery distributions presented here (see sections 6.3 and 8.4.1). In these cases, one would expect the CAM dose to be the preferred value for all distributions, but in fact CAM is the predicted model for only 7 (25%) of the distributions (see Table 4). For the remaining 21 laboratory dose distributions the decision tree predicts that MAM-3, MAM-4 or the lowest 5% (L5%) should be applied. However, these underestimate the given dose by ~50% on average.

8.2 Central age modelling

As discussed above (e.g. section 5), we take the good agreement between multi-grain quartz and $pIRIR_{290}$ multi-grain feldspar ages as evidence that both minerals were in fact effectively well-bleached at deposition. Thus, in the absence of post-depositional mixing (see below) the best estimate of the burial dose should be obtained using the average dose (weighted or unweighted) even if the dose distributions are significantly affected by dose rate inhomogeneity (Guérin et al., 2013). For single-grain dose distributions containing only positive dose estimates CAM is the preferred model; Arnold el al. (2009) argue it is better suited to the statistical properties of such datasets, particularly for older samples. However, for completeness, we have also calculated the CAM_{UL} averages (data not shown) for the single-grain data sets summarised in Table 5. The average ratio of the CAM_{UL} to CAM doses is 0.987±0.005 (n=32); none of the individual ratios are significantly different from unity. Clearly any difference between these two methods of analysis is trivial and the CAM_{UL} data are not considered further.

The ratios of the natural single-grain CAM doses to the demonstrably accurate multi-grain quartz average doses range between 0.83±0.04 and 0.91±0.04 with an average of 0.87±0.02 (n=4; see Table 4). This result seems to imply that the uncertainties used for weighting in CAM are inappropriate. We therefore also calculate a simple arithmetic mean of the accepted doses, discarding the uncertainties assigned to individual dose estimates. This approach results in an average ratio to multi-grain quartz doses of 0.93±0.02 (n=4), which still underestimates the multi-grain data; although by less than the CAM doses (see Table 5).

8.3 Finite mixture modelling

Jacobs et al., 2006; David et al., 2007; Bateman et al., 2007 (among others) also advocate using the Finite Mixture Model (FMM) to dose distributions with a relative OD larger than 20%, with the intention of identifying discrete dose populations within a measured dose distribution. When the FMM identifies more than a single dose component it has generally been assumed that the different populations have arisen because of beta dose heterogeneity and/or post-depositional mixing. Arnold et al. (2009) suggested that even well-bleached dose distributions are log-normally distributed due to inherent error properties and as an extension of this Arnold and Roberts (2011) suggested that the most appropriate statistical model for a given dose distribution could be determined making use of the relative OD as well as the log-weighted skewness. If we assume log-normal distributions, then our four natural single-grain dose distributions are all critically negatively skewed except sample 092204, for which the skew is not critical. Following Arnold and Roberts (2011) we should therefore apply the FMM to the data from 092204 and either the CAM or the FMM[1] to the remaining three data sets. Applying the FMM to the natural dose distributions (where individual uncertainties are based on counting statistics, curve fitting errors and an additional intrinsic uncertainty of 15%), three dose components are identified for all samples except for sample 092202, for which four dose components are identified[2] (see Table 6). In Figure 5 the components identified by the FMM are shown as red lines (see also Figure S5). If we select the most prominent dose components as representing the burial dose of interest (as is usually done in the literature) we obtain FMM to multi-grain dose ratios of 0.65±0.02, 0.84±0.04, 0.88±0.03 and 0.85±0.04 for 092201, -02, -03 and -04, respectively (see Figure 3). Thus, all samples underestimate the MG quartz dose significantly, by >10%. It is not surprising that the FMM results are indistinguishable from the CAM doses for the two younger samples because the most prominent dose component identified by the FMM includes ~75% of the grains and the two bracketing components contain ~half of the remaining grains each (see Table 6). This is in contrast to the results obtained for sample 092201. For this sample, only 52±5% of the grains is identified as part of the most prominent dose component. The FMM identifies a second component containing almost all of the remaining grains (47±5%), which gives a dose of 131±4 Gy (ratio to multi-grain quartz is 1.20±0.05). In the case of sample 092202, the FMM identifies two main components, where the most prominent component (58±8%) is 89±4 Gy and the second component (33±8%) is 142±8 Gy (ratio to multi-grain quartz is 1.33±0.08).

In Figure 6 we present the FMM results as a function of additional uncertainty. For additional uncertainties ranging between 10 and ~20% the burial dose predicted by FMM is approximately constant (as also observed by e.g. Jacobs et al., 2008b) for the two younger samples (092203 and -04). For the two older

[1] In Arnold and Roberts (2011) a single dose distribution is classified as critically negatively skewed and both CAM and FMM ages are listed as preferred ages.

[2] The FMM was run using two to six dose components and the optimum number of components was determined by optimising the Baysian Information Criterion, BIC (Galbraith, 2005).

samples (092201 and -02) the FMM consistently identifies two main components (of similar proportions), which cannot be said to be constant in the range from 10 to 20% as both components decrease when the number of components identified decreases from 4 to 3 between 12.5 and 15% (092201) and 15 and 17.5% (092202). Which of these components that is the most prominent one is very dependent on the additional uncertainty used, e.g. for sample 092201 the most prominent component is ~135 Gy for an additional uncertainty <12.5%, but for uncertainties ranging between 15 and 22.5% the most prominent component is ~71 Gy and for uncertainties larger than 22.5% the most prominent component is ~120 Gy. If we use the BIC and the maximum log likelihood (llik) to optimise the additional uncertainty parameter (Galbraith, 2005), then we should use an additional uncertainty of 25, 22.5, 17.5 and 22% for samples 092201, -02, -03 and -04, respectively. However, doing so does not significantly change the dose estimates of the individual components, except for sample 092202, where the number of components changes from 4 to 3 (see Table 6) and thus the dose estimates of the individual components must change. For both the older samples, optimising the BIC and llik scores results in a shift between the two main dose components such that the most prominent component (50±10% and 57±20 %) for both samples is the older of the two main components, although one could argue that the younger of the two main components is equally likely (49±10% and 43±20 %). The ratios between the two main components remain unaltered by optimisation of the BIC and llik scores. The ratios of the optimised older dose components and the multi-grain dose estimates are 1.13±0.07 and 1.13±0.11 whereas the ratios for the younger components are 0.66±0.05 and 0.72±0.08 for 092201 and -02, respectively. Both of the doses from the older dose components are consistent with the age control provided by ^{14}C and $pIRIR_{290}$ from feldspar (for sample 092202; there is no feldspar result available for 092201). It has also been argued that the different components identified by the FMM could result from beta dose heterogeneity and that the best burial age estimate is obtained using the most prominent dose population divided by an "adjusted" dose rate (e.g. Jacobs et al., 2008a; Jacobs et al., 2008c, Jacobs, 2010; Gliganic et al., 2012). However, Guérin et al. (2013) point out that this approach is unsound, in that it has proved very difficult to simulate even bimodal dose rate distributions using realistic grain size and composition assumptions. These authors conclude that although beta dose rate heterogeneity will produce an increase in the distribution of doses, this distribution should be continuous, and not made up of separable components. Most importantly, these authors also conclude that, if the dose rate to individual grains is not known but only the dose rate distribution, then the best estimate of burial age for well-bleached samples exposed to a heterogeneous beta dose rate is obtained by dividing an average dose by the average dose rate.

Thus, we conclude that if we apply standard analytical approaches to our natural single-grain dose distributions for samples 092203 and -04 we underestimate the burial dose, no matter which statistical model is chosen. This also applies to sample 092201 and -02, except for the FMM result using the optimised additional uncertainty. In these two cases, the largest component is just consistent with the expected burial dose. Nevertheless, we argue later (section 9) that the physical mixing implied by these results is extremely

unlikely on sedimentary grounds and so it would appear that the general assumption that single-grain ages are always superior to multi-grain ages is incorrect.

We have also applied the FMM to the data from the single-grain gamma dose recovery experiments (see section 3.1 and Table 4). In these experiments (both single- and multi-grain) the samples were bleached for 2 hr in the daylight simulator prior to irradiation with ^{60}Co gamma rays. The samples were given gamma doses of 4.9, 9.5, 28.1, 74.8 and 107.4 Gy and subsequently measured using single-grains. We first assigned uncertainties to individual dose estimates based on photon counting statistics, curve fitting errors and an additional uncertainty of 5% (~corresponding to the uncertainty arising from the reproducibility of the instrument). In every case, the FMM identifies multiple components, either 2 or 3 components in each of the 12 distributions (data not shown). The average dose recovery ratio based on the most prominent component is 0.97±0.3 (n=12), indistinguishable from that obtained using CAM (i.e. 0.98±0.02); on average only 58±3% of the accepted dose estimates are assigned to the most prominent component, despite the fact that these irradiations were designed to give a uniform absorbed dose. If we use the BIC and llik scores to optimise both the number of dose components as well as the additional uncertainty value (ranging between 9 and 20%), the FMM predicts that six of the samples should be calculated using CAM (i.e. one component) and that the remaining six samples contain 88±5% of the accepted grains in the most prominent of up to three dose components. The average dose recovery ratio using the main component is 0.98±0.02 (n=12). In the most extreme case (sample 092201 given a gamma dose of 28.1 Gy), the FMM identifies three components: i) 16.8±1.2 Gy (4±2%), ii) 26.6±0.5 Gy (66±6%) and iii) 36.8±1.2 Gy (30±6%) with an additional uncertainty of 9% following optimisation of the BIC and llik scores. Even omitting this extreme case, the FMM predicts that only 92±2% of the accepted grains are contained in the most prominent component. We deduce that, in this well-controlled experiment, the FMM identifies what can best be described as phantom dose components in our uniformly irradiated quartz dose recovery distributions. The components identified here by the FMM must be related to intrinsic luminescence properties, rather than extrinsic phenomena such as mixing.

8.4 Additional/alternative single-grain rejection criteria

For the multi-grain natural measurements only ~1% of the aliquots were discarded because of apparent saturation, but for the natural single-grain measurements ~22 and 11% of the single grains were discarded for the two older (092201 and -02) and the two younger (092203 and -04) samples, respectively (see section 3.3 and Table 5). Discarding grains in saturation has the potential to bias the resulting dose distributions and this is considered below.

8.4.1 Dose recovery

The gamma dose recovery experiments (both single- and multi-grain) were undertaken to assess the suitability of the chosen SAR protocol as well as the variability arising from intrinsic sources of uncertainty in our single-grain dose distributions. When grains are irradiated using a gamma source, the energy deposition can readily be arranged to be effectively uniform; this is not the case when irradiated using a beta source. Thus in a gamma dose recovery experiment, one is testing the ability of the instrumentation and measurement protocol to measure a dose which is known to be the same in all grains; any dispersion in the data not accounted for by known measurement uncertainties (counting statistics, fitting, instrument reproducibility, etc.) must arise because of inaccuracies arising from the measurement protocol employed. The results from the gamma dose recovery experiments are summarised in Table S1 (multi-grain) and Table 4 (single-grains). As stated in section 6.3, all dose recovery ratios are acceptable (i.e. within ±10% of unity if uncertainties are taken into account) although there is a tendency for the dose recovery ratio to decrease as a function of given dose for doses larger than 30 Gy, except for sample 092202. This trend is also observed in multi-grain measurements. A similar result was reported for both single-grain and multi-grain dose recovery experiments by Thomsen et al. (2012).

Gliganic et al. (2012) observed a beta dose recovery ratio of 0.75±0.06 (OD=27±3%) for their sample, when a dose of 121 Gy was given. This underestimate was attributed to the inclusion of grains which either had D_0 values less than 25 Gy or to a decreasing sensitivity >30%. By removing such grains the authors obtained a beta dose recovery ratio of 0.92±0.04 (OD=18±6%). If we apply the same additional rejection criteria to our 107.4 Gy gamma dose recovery distribution for sample 092201, we reject four grains because of the sensitivity change criterion and two grains because of the D_0 criterion and observe that the dose recovery ratio changes from 0.88±0.03 (n=119) to 0.90±0.03 (n=113) and the OD from 29±3% to 25±2%. Although this change cannot be argued to be significant we note that the small apparent increase in dose recovery ratio comes from the removal of the two grains with D_0 values less than 25 Gy. Figure 7 shows the D_0 distribution for the portion of sample 092201 given a gamma dose of 107.4 Gy. The D_0 values range between 19±7 and 455±157 Gy with a weighted average value of 105±6 Gy (n=119). Although little is currently known about the accuracy of dose estimation in quartz near the saturation limit, it seems unlikely that a grain with a D_0 value of 25 Gy should be able to record a given dose of 107 Gy (at ~4×D_0) accurately. If we arbitrarily assume that the D_0 value should be larger than half the dose to be measured (as suggested by Wintle and Murray, 2006) then for this data set ~10% of the grains have D_0 values smaller than 53 Gy (half of the given dose). The reliability of all these dosimeters can be questioned, even although they pass the standard single grain rejection criteria given in section 3.3.

In the inset to Figure 7 we show the effect of rejecting individual grains based only on their D_0 values on the dose recovery ratio for both the arithmetic mean (squares) and the CAM dose (circles). For x=0 all grains are included in the dose recovery ratio estimation, but for x=100, only grains with D_0>100 Gy are

included. The dose recovery ratio increases with increasing D_0 value until $D_0 \geq 100$ Gy, where after the dose recovery ratio does not change significantly. The dose recovery ratio for $D_0 \geq 100$ Gy is 1.02±0.03 and 0.99±0.02 (n=76) for the arithmetic mean and the CAM dose, respectively. Thus, from this data set it would appear that grains should have D_0 values approximately ≥ the dose to be measured before they are likely to act, on average, as accurate dosimeters. A similar analysis was done for the other gamma or beta irradiated portions and the results are given in Table 4. The apparent dependency of dose recovery ratio on dose is eliminated if grains with D_0 values less than the given dose are rejected. Applying the same criterion to the data sets obtained for the gamma irradiated portions of sample 092202 does not change the dose recovery ratio significantly (i.e. it is still acceptable).

If the D_0 value of individual grains is not taken into consideration then the relative OD appears to increase with given dose, e.g. the relative average OD for the sample given 4.9 Gy is 15±2% and increases with dose to 29±3% for the sample given a dose of 107.4 Gy (similar to the observation by Thomsen et al. 2012). The average relative OD for all gamma dose recovery experiments is 19.0±1.1 % (n=12). When the D_0 value criterion is taken into consideration this apparent dependence on given dose disappears, and the average relative OD is 16.0±0.8 % (n=12 samples, Table 4). In the literature (e.g. Thomsen et al., 2005; 2012; Sim et al., 2013) it has been shown that the relative OD observed in single-grain gamma dose recovery experiments tends to be larger than that observed in beta dose recovery experiments. For the beta dose recovery experiments here, the average relative OD is 17.7±1.5% (n=16 samples, Table 4) before application of the D_0 criterion and 13.0±0.9% after.

Thus, applying the D_0 selection criterion improves the dose recovery ratios for doses >30 Gy and reduces the apparent relative ODs. However, for the samples given gamma doses of 74.8 and 107.4 Gy this criterion also results in the rejection of 34±6 % (n=6 samples) of the otherwise accepted grain population.

Thomsen et al. (2012) showed for their gamma dose recovery experiment at 208 Gy that the grain population with naturals in saturation had an average D_0 value lower than that of the accepted grain population. Thus, one would expect that the number of grains in saturation will be reduced significantly by applying the D_0 selection criteria. For the four gamma dose distributions given doses similar to the natural dose (e.g. ~107 Gy for 092201 and -02 and ~75 Gy for 092203 and -04) 170 grains (~25%) of the otherwise 704 accepted grains were rejected because they were deemed to be in saturation. After application of the D_0 selection criteria only 3 grains (1%) of the remaining 330 grains appear to be in saturation and the D_0 values of these grains are all relatively poorly known (s_{rel} ~55%) and close to the D_0 selection value (i.e. the average ratio of D_0 to D_0 selection value is 1.09±0.01, n=3) . Thus, it would appear that any bias likely to occur because of a significant number of individual grains appearing to be in saturation is overcome by the application of the D_0 selection criteria.

8.4.2 Single-grain burial doses with D_0 selection

The effect on the estimated burial dose (arithmetic mean or CAM) of only accepting grains with a D_0 value above a certain threshold value is shown for the four natural single-grain dose distributions in Figure 8. The effect of eliminating grains with low D_0 values is most clearly shown for sample 092201, where the burial dose increases by ~15%, when grains with D_0 values < 100 Gy are rejected. If the D_0 criterion is tightened further there is no significant effect on the estimated burial dose. For sample 092202 there is no significant change in burial dose as grains with lower values of D_0 are rejected. For samples 092203 and -04, there is a systematic increase in burial dose of ~4%, when grains with D_0 values < 50 Gy are rejected. If we apply the D_0 criterion outlined in section 8.4.1 (i.e. the minimum D_0 value must be equal to the estimated burial dose), we obtain minimum D_0 values of ~110, 100, 65 and 65 Gy; the number of grains in the accepted grain populations are reduced by 50, 45, 27 and 25% for samples 092201, -02, -03 and -04, respectively. If we calculate the arithmetic mean of the remaining dose populations, all individual ratios (with respect to the multi-grain quartz results) are consistent with unity and the average ratio is 1.00±0.04 (n=4 samples). If we apply the CAM then all ratios are systematically lower than unity with an average ratio of 0.93±0.03. The average relative OD decreases from 33.6±1.4 to 27.5±0.5% (see Table 5) and the average number of accepted grains decreases by 37±7%. Similar results are obtained if we apply the D_0 criterion to the dose distributions resulting from only applying the s_{Tn}<20% rejection criterion, i.e. recycling, IR depletion and recuperation are not taken into consideration (see Table 5).

8.4.3 Single-grain burial doses with F_R selection

It is well-documented that the decay shapes of individual grains of quartz are highly variable (e.g. Duller, 2008 and references therein). To illustrate the variability observed in signals from our samples, we show representative regenerated stimulation curves for sample 092201 in the inset to Figure 4d. In most single-grain studies the ubiquitously observed differences in stimulation curves are not taken into consideration in the data analysis. However, Duller (2012) showed in his single-grain beta dose recovery experiments that an apparent dose underestimation at high doses was removed by only including the fastest decaying signals in the data analysis, i.e. by only including the grains with the largest "Fast ratio, F_R" values (Madsen et al., 2009; Durcan and Duller, 2011) with F_R given by (*F-BG*)/(*M-BG*), where *F* is the sum of the initial 20 ms, *M* the following 40 ms and *BG* the last 150 ms of stimulation. Unfortunately, this approach is considerably complicated by the variation in effective stimulation power (Thomsen et al., 2012; 2015; submitted) and as a result many grains are likely to be rejected because of poor laser-grain coupling rather than true differences in decay rate. However, although only choosing the grains with the largest F_R values is very expensive in terms of data reduction, Thomsen et al. (submitted) nevertheless show that the ratio of measured to expected dose increases systematically for these samples, when only grains with F_R>4 are included. They show that the average ratio of single-grain CAM dose to multi-grain arithmetic dose increases from 0.87±0.02

(standard rejection criteria) to 0.95±0.02 (n=4 samples, F_R>4). The results of Thomsen et al. (submitted) are given in Table 5 for completeness. Again, if we apply this rejection criterion to the dose distributions resulting from only applying the s_{Tn}<20% rejection criterion, the results are indistinguishable from those obtained when applying the F_R criterion subsequent to standard rejection criteria.

8.4.4 Single-grain burial doses using the two additional/alternative rejection criteria

In the previous sections (8.4.2 and 8.4.3) we showed that by considering the curvature of individual single-grain DRCs (D_0) and the decay rate (F_R) we obtain significant improvements in the apparent accuracy of the single-grain burial doses, i.e. single-grain CAM to multi-grain dose ratios increase on average from 0.87±0.02 to 0.93±0.03 (D_0 criterion) and 0.95±0.02 (F_R criterion). Since there appears to be no correlation between the grains being selected by these two additional or alternative rejection criteria (i.e. grains with a large D_0 can have a small F_R value and vice versa), we apply both these additional rejection criteria to our natural dose distributions. The open symbols in Figure 5 represent the dose estimates rejected using both these alternative rejection criteria, and the closed symbols represent the remaining dose estimates. The resulting doses are summarised in Table 5 and Figure 9. The average relative number of accepted grains for the four natural dose distributions has been reduced by an order of magnitude from ~8% (standard rejection criteria) to 0.8%, but the average measured to expected dose has increased from 0.87±0.02 to 1.04±0.02 (CAM) and from 0.93±0.02 to 1.07±0.03 (average). The average relative OD has been reduced by ~50%, even though there is still significant unexplained variability in the resulting dose distributions (18±1% on average. Note that the resulting dose distribution for sample 092203 is under-dispersed, probably by chance since it only contains 9 grains, which is generally considered to be too few for burial dose estimation). Very similar results are obtained if we use the D_0 and F_R rejection criteria on the dose distributions obtained by only applying the s_{Tn}<20% rejection criteria (see Table 5), i.e. the average measured to expected dose increases from 0.87±0.02 to 1.02±0.02 (CAM) and 0.93±0.03 to 1.05±0.03 (average).

9 Summary and discussion

Single-grain data are often thought to be superior to multi-grain data because it is argued (e.g. Arnold and Roberts, 2011; Jacobs et al., 2012) that individual grains with aberrant OSL characteristics can be discarded by appropriate rejection criteria. Here we have tested the impact of these standard rejection criteria using both multi-grain and single grain quartz measurements. In contrast to the observation by Jacobs et al. (2006) we do not find that the application of these rejection criteria to our samples changes either the dose or the dispersion significantly.

Particularly in the single-grain literature, it is common to provide tables detailing how many grains failed each rejection criterion - in our opinion, it would be more informative to provide documentation of the

effect on D_e and OD of applying these criteria. It is of course true that e.g. a poor recycling ratio is indicative of inaccurate dose estimation, but one must ask whether the possible inaccuracies in the resulting dose estimates bias the mean or contribute significantly to the dispersion in the data? If not, then the effect of these possible inaccuracies must be negligible compared to other sources of dispersion, and in this case discarding these data does not strengthen the overall data set and indeed may weaken it. In our view, if the only observable effect of applying these rejection criteria is to reduce the number of accepted grains, there is no justification for applying them.

To estimate burial doses for the multi-grain dose distributions we use the simple arithmetic average (i.e. the individual uncertainties are not taken into account). The ratio of the CAM dose to the unweighted average is 0.981±0.004 (n=4 samples); all four ratios are individually consistent with unity. Thus, we do not observe any justification for weighting the individual aliquots according to their individual precisions based on counting statistics (and curve fitting). The calculated over-dispersions are consistent with the relative standard deviations and thus there is, as expected, no need for more sophisticated modelling.

To estimate the burial doses from single-grain distributions we have used several different approaches. If we follow the recommendations of both Olley et al. (2004b) and Bailey and Arnold (2006) we should use a minimum age model to obtain accurate burial dose estimates. Had these data been analysed without any prior knowledge as to the expected dose this would have resulted in an underestimation of the true ages by more than 40%. Similar results have been reported for known age modern (40-500 years) palaeoflood deposits (Medialdea et al., 2014). In addition, if we apply the decision tree of Bailey and Arnold (2006) to the dose recovery distributions (both gamma and beta) then we find that ~75% (21 out of 28) require minimum age modelling. Since these doses were arranged to be absorbed uniformly by all grains (especially in the case of the gamma irradiations), this is obviously inappropriate. Thus, for these samples at least, we conclude that the OD and/or the weighted skewness and kurtosis should not be used to decide whether minimum age modelling is required. Since it appears that one cannot reliably determine from internal evidence whether or not a single grain distribution is incompletely bleached, we argue that minimum age modelling should only be applied where there is independent evidence that significant incomplete bleaching is present. This evidence should obviously be as non-subjective as possible, and be based on e.g. comparison with independent age control, analysis of modern analogues or comparison of quartz and feldspar ages. Subjective interpretation of contextual evidence is insufficient to justify the application of particular statistical models.

The FMM is widely used in the single-grain literature to test for the existence of individual dose components in dose distributions; when several dose components are identified they are usually explained in terms of beta dose rate heterogeneity or post-depositional mixing. However, as Guérin et al. (2013) point out, beta dose rate heterogeneity is most unlikely to generate discrete dose populations, and in any case the best

estimate of burial dose for well-bleached, unmixed samples exposed to a heterogeneous beta dose rate is obtained from the average dose and the average dose rate. Only when we are able to generate a burial dose rate appropriate to a particular individual grain for which we have a dose measurement will this conclusion be changed.

In this study, if the assigned uncertainties are based on photon counting statistics, curve fitting errors and an additional uncertainty σ_{add} of 15% (same approach as that of e.g. Arnold and Roberts, 2011) then the FMM identifies three or four components in each of our four natural dose distributions (see Table 6). These components are similar to each other both in terms of dose and proportion of grains for the two younger samples (092203 and -04). Given that this pair of samples was taken from the same unit (i.e. FDG F) it could be argued that this is to be expected, and thus that this is an indication of the reliability of the FMM. However, this unit was capped by rock fall from a subsequent roof collapse (see Figure 1) and it is unlikely in the extreme that these samples could have been contaminated by a significant fraction of younger grains. Despite this, both these samples contain a dose component of ~32 Gy, each containing ~15% of the accepted grain population; this component would appear to be ~12 ka younger than the expected age for this unit, but it cannot have been derived from an overlying layer.

In the older sample 092201 the FMM identifies two very different dose components (i.e. ~71 Gy and ~130 Gy) in almost equal proportions (~50% each). In sample 092202 the FMM also identifies two dose components, of ~90 Gy (~60%) and ~140 Gy (~33%). If we optimise the additional uncertainty parameter σ by using BIC and llik scores both 092201 and -02 contain three components and we do not observe significant changes in the doses or their proportions, i.e. we obtain two components of ~74 and ~122 Gy in statistically equal amounts (see Table 6). The older of these components is in fact consistent with the age control, but for these results to be interpreted in terms of post-depositional mixing one would have to be prepared to accept gross vertical mixing of two distinctly different sediment bodies after deposition of the younger, over a vertical distance of ~1 m and without significant input from the intermediate layer. The only mechanism that could move such large quantities of sediment over such large distances is burrowing activity, but to achieve 50% mixing any sedimentary structure would have been completely destroyed. Not only is there no evidence for such burrowing, but stratification is in fact well preserved.

In section 8.4.2 we examined the effect of excluding grains with low D_0 values and find that if $D_0 > D_e$ (sample average) a significant improvement in the average doses is obtained. If the FMM is applied to these selected dose distributions, we find that it continues to identify components very similar to those identified following the use of standard rejection criteria (see Table 6). The ratio of the highest dose component to the 2nd highest dose component is 1.77±0.08 for all four samples regardless of whether or not σ_{add} is optimised. A similar ratio (1.80±0.07, n=8) can be observed from the data presented in Table 1 and 2

in Jacobs et al. (2012) for the samples where the FMM identifies three components and the OD ranges between 29±2 and 47±4 %, i.e. similar to the OD values for the samples presented here. (In Jacobs et al., 2012 the FMM also identifies three components for sample EM08-11, and the ratio between the highest and 2nd highest dose components is 3.4±0.3, but this sample has a much larger OD of 89±3 %.) This markedly consistency between two completely independent data sets may suggest that the FMM identifies clustering in some intrinsic quartz luminescence characteristic (e.g. first SAR cycle sensitivity change) rather than some extrinsic phenomenon such as post-depositional mixing. Unfortunately, this suggestion cannot be tested further because most authors do not provide sufficient detail to allow this calculation.

We also applied the FMM to the single-grain gamma dose recovery distributions and found that only 92% of the dose estimates are contained in the most prominent dose component. Thus, when the FMM is applied to artificially uniformly irradiated material, it identifies dose components that should not exist and when applied to natural dose distributions it identifies components that are completely inconsistent with the known deposition environment. We conclude that, in our study, the components identified by the FMM are not, in any way, related to post-depositional mixing.

As discussed earlier, there is good agreement between ^{14}C ages and standard OSL ages based on multi-grain aliquots (both quartz and feldspar) as reported in Aubry et al. (2012). Furthermore, the good agreement between multi-grain quartz and multi-grain pIRIR$_{290}$ K-rich feldspar ages (the pIRIR$_{290}$ ages are on average 3±1 ka older than the quartz ages) strongly suggests that the samples were well-bleached at deposition since the pIRIR$_{290}$ signal bleaches ~2 orders of magnitude slower than quartz (see e.g. Buylaert et al., 2012, Murray et al., 2012); certainly there is little doubt that the quartz OSL signals were well-bleached at deposition. Bearing in mind that only the average dose rate is available, the best estimator of the single-grain burial dose for these samples should be the average dose. As discussed in section 5, we have chosen for simplicity to compare our single-grain doses to those based on multi-grain quartz. From Figure 3 it is clear that the multi-grain quartz doses are very unlikely to be overestimates and unlikely to be significant underestimates, although this possibility cannot be completely ruled out. If we now compare the single-grain data to the preferred multi-grain quartz data then the single-grain CAM dose (weighted dose) systematically underestimates by ~15%. A similar observation was made by Arnold and Roberts (2011). This could imply either that the weighting in CAM is inappropriate or that the assigned uncertainties are inappropriate. If we use the unweighted arithmetic mean to calculate the burial dose then the single-grain doses underestimate the multi-grain doses by ~7%, i.e. the ratio of measured to expected dose is 0.93±0.02. However in this calculation, a significant proportion of grains were discarded because they were in apparent saturation (~22% and ~11% for the older and younger samples, respectively) and this is very likely to bias the average dose to lower doses.

If we assume that the L_n/T_n values measured from young well-bleached sensitive multi-grain aliquots are normally distributed (i.e. counting statistics are negligible compared to other sources of dispersion), then so should be the dose distribution obtained by interpolating these L_n/T_n values onto the linear part of the dose response curve. In this case the most accurate dose estimate is given by the average dose. As the burial dose increases (and with it the average measured L_n/T_n value) the derived dose distribution will become more and more positively skewed as the L_n/T_n values are interpolated onto the curving part of the dose response curve. Then the median dose value is more accurate than the average dose, as argued by Murray and Funder (2003). In the extreme case where the sample is in true saturation (i.e. the average L_n/T_n value is equal to the saturation value of the dose response curve) then 50% of the measured L_n/T_n values should lie above saturation (and so cannot be interpolated to give a dose); the remaining 50% will give a highly positively skewed dose distribution, the average of which will significantly underestimate the true dose. If individual aliquots are only accepted if $D_e<2\times D_0$ then this systematic error is increased and both the mean and median of the resulting dose distribution is bound to grossly underestimate the true burial dose. Thus, the $D_e<2\times D_0$ threshold (Wintle and Murray, 2006) should not be applied on an aliquot by aliquot basis but rather on a sample by sample basis. In general, whenever a significant proportion of aliquots appear to be in saturation it is likely that a dose calculated on the remaining part of the distribution will be inaccurate.

Based on laboratory dose recovery experiments it was shown in section 8.4.1 that if we only accept grains with a D_0 value larger than the average sample dose we reduce the average grain population in saturation from ~25% to ~1% and remove the apparent dose underestimate as well as the OD dependency on given dose. When we apply the same criteria to the natural single-grain dose distributions, we reduce the relative number of grains in saturation from ~16% to ~3% (11 grains; standard rejection criteria). Of these remaining 11 grains, 2 grains have D_0 values indistinguishable from the relevant value used in the rejection criteria (i.e. the absolute value ignoring the assigned uncertainties remained greater); 6 grains have L_n/T_n values significantly above the saturation value of the saturating exponential, presumably indicating a failure of the measurement protocol for these grains. Thus, for these samples, by imposing an additional rejection criterion involving the D_0 value of individual grains, we are able to produce data sets which should be largely unbiased by the presence of saturated grains. The ratio of the single-grain burial dose to the multi-grain burial dose is then 0.93±0.03 for the CAM and 1.00±0.04 for the arithmetic mean. However, applying this rejection criterion reduces the accepted grain population significantly, by ~48% and ~26% on average for the old and young samples, respectively.

If we accept only the grains with the fastest decaying signals (i.e. $F_R>4$), we see a systematic increase in dose compared to the dose determined using all grains irrespective of decay rate for the natural samples (see Table 5 and Figure 9). However, this reduces the accepted grain population by ~87% and thus it is unlikely that it will be practical to apply this criterion to the majority of samples.

Applying both the D_0 and the F_R rejection criteria reduce the accepted grain population by ~92% and increase the measured to expected dose to 1.07±0.03 (average) and 1.04±0.02 (CAM). Thus, for the arithmetic average we do not observe an improvement in accuracy when applying both alternative rejection criteria (the measured to expected ratio is 0.93±0.02 when only standard rejection criteria are employed). However, a significant improvement in accuracy is observed for CAM; the measured to expected ratio is increased from 0.87±0.02 to 1.04±0.02, but this improvement is only obtained by rejecting ~90% of the detectable grain population with the corresponding inevitable loss of precision on individual dose estimates (by about a factor of 3, see Table 5).

10 Conclusions

In this paper we have applied standard single-grain dating techniques to four archaeological samples. The ages of these samples have been determined by comparison with a previously established ^{14}C chrono-stratigraphy, and by standard multi-grain quartz and K-rich feldspar ($pIRIR_{290}$) OSL measurements; all these methods give consistent ages. From this we conclude that the quartz grains were, on average, well-bleached at deposition.

We find that application of standard single-grain rejection criteria does not result in significant changes in either dose or over-dispersion of the single-grain dose distributions, and conclude that there is no justification for the application of these criteria at least to these samples. Since this work was completed, similar conclusions have been drawn in several other papers (e.g. Geach et al., 2015; Gúerin et al., 2015; Hansen et al., 2015; Kristensen et al., 2015; Zhao et al., 2015) for different samples of different origins. Although the CAM ages derived from these populations underestimate the expected ages, application of either the over-dispersion threshold criterion of 20% (Olley et al., 2004b), or the decision tree suggested by Bailey and Arnold (2006) indicates that minimum age modelling is required to obtain the most accurate estimate of burial age. Such minimum age modelling results in significant age underestimation, of more than 40% (depending on which model is used). We have also applied the FMM and find that for the two young samples (~20 ka) a statistically significant but physically unrealistic young dose component is identified. The most prominent dose component (~80%) for these samples underestimates the multi-grain quartz dose by more than 10%. For the two older samples the FMM identifies two equally likely dose components. Although in one case the older of these components does result in an age that is just consistent with the radiocarbon age range (although older than the feldspar ages), the physical mixing implied by these results is rejected as physically unrealistic. It is deduced that the components identified by the FMM do not relate to extrinsic processes such as incomplete bleaching or physical mixing, but may rather reflect intrinsic luminescence characteristics.

It is often argued that single-grain data are superior to multi-grain data, because the multi-grain data must contain signals from 'inappropriate' grains (e.g. grains in saturation and grains failing the recycling test). However, for these samples at least, it would appear that single-grain quartz results are in poorer agreement with the age control than multi-grain data if only standard rejection criteria are applied. Single-grain CAM data do become acceptable if more stringent rejection criteria are introduced (based on D_0 and F_R), but this is at the expense of a considerable reduction in the number of accepted grains and thus a decrease in precision by about a factor of 3. This conclusion applies whether or not standard rejection criteria are used before these more stringent criteria. Given the very large fraction of single-grains that must be rejected to provide accurate single-grain dose estimates, it is of course surprising that multi-grain dose estimates (based on the sum of signals from acceptable as well as unacceptable individual grains) provide accurate dose estimates without any further data selection.

Finally, we conclude that, for these quartz samples, with doses in the range ~60 to 110 Gy, multi-grain analysis gives the most accurate ages; there is no advantage in the use of single grains. Indeed, had these samples been analysed in the absence of other age control, the application of standard single-grain methods would have led to significant misinterpretation of results and a corresponding inaccuracy in ages.

References

Aitken, M.J., 1985. Thermoluminescence Dating. Academic Press, London.

Aitken, M.J., 1998. An Introduction to Optical Dating. Oxford University Press, Oxford.

Armitage, S.J., Jasmin, S.A., Marks, A.E., Parker, A.G., Usik, V.I., Uerpmann, H., 2011. The Southern Route "Out of Africa": Evidence for an Early Expansion of Modern Humans into Arabia. Science 331, 453-456.

Arnold, L. J., Demuro, M., Navazo, M., Benito-Calvo, A., Pérez-González, A., 2013. OSL dating of the Middle Palaeolithic Hotel California site, Sierra de Atapuerca, north-central Spain. Boreas 42, 285–305.

Arnold, L., Roberts, R.G., 2011. Paper I – Optically stimulated luminescence (OSL) dating of perennially frozen deposits in north-central Siberia: OSL characteristics of quartz grains and methodological considerations regarding their suitability for dating. Boreas 40, 389–416.

Arnold, L.J., Roberts, R.G., 2009. Stochastic modelling of multi-grain equivalent dose (De) distributions: Implications for OSL dating of sediment mixtures. Quaternary Geochronology 4, 204–230.

Arnold L.J., Roberts, R.G., Galbraith, R.F., DeLong, S.B., 2009. A revised burial dose estimation procedure for optical dating of young and modern-age sediments. Quaternary Geochronology 4, 306–325.

Aubry, T., Dimuccio.L.A., Buylaert, J.P., Liard, M., Murray, A.S, Thomsen, K.J., Walter, B., 2014. Middle-to-Upper Palaeolithic site formation processes at the Bordes-Fitte rockshelter (Central France). Journal of Archaeological Science 52, 436–457.

Aubry, T., Dimuccio.L.A., Almeida, M., Buylaert, J.P., Fontana, L., Higham, T., Liard, M., Murray, A.S, Neves, M.J., Peyrouse, J-B., Walter, B., 2012. Stratigraphic and technological evidence from the middle palaeolithic-Châtelperronian-Aurignacian record at the Bordes-Fitte rockshelter (Roches d'Abilly site, Central France). Journal of Human Evolution 62, 116-137.

Bailey, R.M., Arnold, L.J., 2006. Statistical modelling of single grain quartz De distributions and an assessment of procedures for estimating burial dose. Quaternary Science Reviews 25, 2475–2502.

Banerjee, D., Murray, A.S., Bøtter-Jensen, L., Lang, A., 2001. Equivalent dose estimation using a single aliquot of polymineral fine grains. Radiation Measurements 33, 73–94.

Bateman, M.D., Boulter, C.H., Carr, A.S., Frederick, C.D., Peter, D., Wilder, M., 2007. Detecting post-depositional sediment disturbance in sandy deposits using optical luminescence. Quaternary Geochronology 2, 57–64.

Buylaert, J.P., Jain, M., Murray, A. S., Thomsen, K. J., Thiel, C., Sohbati, R., 2012. A robust feldspar luminescence dating method for Middle and Late Pleistocene sediments. Boreas 41, 435–451.

Bordes, F., Fitte, P., 1950. Un abri solutréen à Abilly (Indre et Loire). Bull. Soc. Préhist. Fr. 47, 146-153.

Bronk Ramsey, C., 2013. OxCal 4.2. http://c14.arch.ox.ac.uk/oxcal

Bøtter-Jensen, L., Thomsen, K.J., Jain, M., 2010. Review of optically stimulated luminescence (OSL) instrumental developments for retrospective dosimetry. Radiation Measurements 45, 253-257.

Bøtter-Jensen, L., Andersen, C.E., Duller, G.A.T., Murray, A. S., 2003. Developments in radiation, stimulation and observation facilities in luminescence measurements. Radiation Measurements 37, 535-541.

Choi, J.H., Murray, A.S., Cheong, C.-S., Hong, S.C., 2009. The dependence of dose recovery experiments on the bleaching of natural quartz OSL using different light sources. Radiation Measurements 44, 600–605.

David, B., Roberts, R.G., Magee, J., Mialanes, J., Turney, C., Bird, M., White, C., Fifield, K., Tibby, J., 2007. Sediment mixing at Nonda Rock: investigations of stratigraphic integrity at an early archaeological

site in northern Australia and implications for the human colonisation of the continent. Journal of Quaternary Science 22, 449–479.

Demuro, M., Froese, D.G., Arnold, L.J., Roberts, R.G., 2012. Single-grain OSL dating of glaciofluvial quartz constrains Reid glaciation in NW Canada to MIS 6. Quaternary Research 77, 305–316.

Duller, G.A.T., 2012. Improving the accuracy and precision of equivalent doses determined using the optically stimulated luminescence signal from single grains of quartz. Radiation Measurements 47, 770–777.

Duller, G.A.T., 2008. Single-grain optical dating of Quaternary sediments: why aliquot size matters in luminescence dating. Boreas 37, 589–612.

Duller, G.A.T., 2007. Assessing the error on equivalent dose estimates derived from single aliquot regenerative dose measurements. Ancient TL 25, 15-24.

Duller, G.A.T., 2003. Distinguishing quartz and feldspar in single grain luminescence measurements. Radiation Measurements 37, 161–165.

Durcan, J.A., Duller, G.A.T., 2011. The fast ratio: A rapid measure for testing the dominance of the fast component in the initial OSL signal from quartz. Radiation Measurements 46, 1065-1072.

Feathers, J.K., Rhodes, E., Huot, S., McAvoy, J., 2006. Luminescence dating of sanddeposits related to late Pleistocene human occupation at the Cactus Hill site, Virginia, USA. Quaternary Geochronology 1, 167-187.

Feathers, J.K., 2003. Single-grain OSL dating of sediments from the Southern High Plains, USA. Quaternary Science Reviews 22, 1035–1042.

Galbraith, R.F., 2005. Statistics for Fission Track Analysis. Chapman & Hall, London.

Galbraith, R., Roberts, R.G., Laslette, G., Yoshidha, Olley, J., 1999. Optical dating of single and multiple grain Quartz from Jinmium Rock Shelter, Northern Australia. Part I, Experimental design and statistical models. Archaeometry 41, 339-364.

Galbraith, R.F., Green, P.F., 1990. Estimating the component ages in a finite mixture. Nuclear Tracks and Radiation Measurements 17, 197-206.

Geach, MR; Thomsen, KJ; Buylaert, J-P; Murray, AS; Mather, AE; Telfer, MW; Stokes, M; 2015. Single-grain and multi-grain OSL dating of river terrace sediments in the Tabernas Basin, SE Spain. Quaternary Geochronology (in press).

Gliganic, L.A., Jacobs, Z., Roberts, R.G., 2012. Luminescence characteristics and dose distributions for quartz and feldspar grains from Mumba rockshelter, Tanzania. Archaeological and Anthropological Sciences 4, 115–135.

Guérin, G., Frouin, M., Talamo, S., Aldeias, V., Bruxelles, L., Chiotti, L., Dibble, H.L., Goldberg, P., Hublin, J.-J., Jain, M., Lahaye, C., Madelaine, S., Maureille, B., McPherron, S.P., Mercier, N., Murray, A.S., Sandgathe, D., Steele, T.E., Thomsen, K.J., Turq, A., 2015. A multi-method luminescence dating of the Palaeolithic Sequence of La Ferrassie based on new excavations adjacent to the La Ferrassie 1 and 2 Skeletons. Journal of Archaeological Science 58, 147-166.

Guérin, G., Murray, A.S., Jain, M., Thomsen, K.J., Mercier, N., 2013. How confident are we in the chronology of the transition between Howieson's Poort and Still Bay? Journal of Human Evolution 64, 314-317.

Guérin, G., Mercier, N., Adamiec, G., 2011. Dose-rate conversion factors: update. Ancient TL 29, 5-8.

Hansen, V., Murray, A.S., Buylaert, J.-P., Yeo, E.-Y., Thomsen, K.J., 2015. A new irradiated quartz for beta source calibration. Radiation Measurements (in press).

Higham, T.F.G., Jacobi, R.M., Bronk Ramsey, C., 2006. AMS Radiocarbon dating of ancient bone using ultrafiltration. Radiocarbon 48, 179–195

Jacobi, R.M., Higham, T.F.G., Bronk Ramsey, C., 2006. AMS radiocarbon dating of Middle and Upper Palaeolithic bone in the British Isles: improved reliability using ultrafiltration. Journal of Quaternary Science 21, 557–573.

Jacobs, Z., Roberts, R.G., Nespoulet, R., El Hajraoui, M.A., Debénath, A., 2012. Single-grain OSL chronologies for Middle Palaeolithic deposits at El Mnasra and El Harhoura 2, Morocco: Implications for Late Pleistocene huma-environment interactions along the Atlantic coast of northwest Africa. Journal of Human Evolution 62, 377-394.

Jacobs, Z., Meyer, M.C., Roberts, R.G., Aldeiasc, V., Dibble, H., El Hajraoui, M.A., 2011. Single-grain OSL dating at La Grotte des Contrebandiers ('Smugglers' Cave'), Morocco: improved age constraints for the Middle Paleolithic levels. Journal of Archaeological Science 38, 3631-3643.

Jacobs, Z., 2010. An OSL chronology for the sedimentary deposits from Pinnacle Point Cave 13B-A punctuated presence. Journal of Human Evolution 59, 289-305.

Jacobs, Z., Roberts, R.G., Galbraith, R.F., Deacon, H.J., Grün, R., Mackay, A., Mitchell, P., Vogelsang, R., Wadley, L., 2008a. Ages for the Middle Stone Age ofsouthern Africa: implications for human behavior and dispersal. Science 322, 733-735.

Jacobs, Z., Wintle, A.G., Roberts, R.G., Duller, G.A.T., 2008b. Equivalent dose distributionsfrom single grains of quartz at Sibudu, South Africa: context causes and consequences for optical dating ofarchaeological deposits. Journal of Archaeological Science 35, 1808-1820.

Jacobs, Z., Wintle, A.G., G.A.T., Roberts, R.G., Duller, Wadley, L., 2008c. New ages for the post-Howiesons Poort, late and final Middle Stone Age at Sibudu, South Africa. Journal of Archaeological Science 35, 1790-1807.

Jacobs, Z., Duller, G.A.T., Wintle, A.G., 2006. Interpretation of single grain De distributions and calculation of De. Radiation Measurements 41, 264-277.

Jacobs, Z., Duller, G.A.T., Wintle, A.G., 2003. Optical dating of dune sand from Blombos Cave, South Africa: II - single grain data. Journal of Human Evolution 44, 613-625.

Jain, M., Murray, A.S., Bøtter-Jensen, L., Wintle, A.G., 2005. A single-aliquot regenerative-dose method based on IR (1.49 eV) bleaching of the fast OSL component in quartz. Radiation Measurements 39, 309-318.

Kristensen, J.A., Thomsen, K.J., Murray, A.S., Buylaert, J.P., Jain, M., Breuning-Madsen, H., 2015. Quantification of termite bioturbation in a savannah ecosystem: application of OSL dating. Quaternary Geochronology (in press).

Lapp, T., Jain, M., Thomsen, K.J., Murray, A.S., Buylaert, J.P., 2012. New luminescence measurement facilities in retrospective dosimetry. Radiation Measurements 47, 803-808.

Madsen, A.T., Duller, G.A.T., Donnelly, J.P., Roberts, H.M., Wintle, A.G. (2009). A chronology of hurricane landfalls at Little Sippewisset Marsh, Massachusetts, USA, using optical dating. Geomorphology 109, 36-45.

Marquet, J.C., 2011. La Préhistoire en Touraine. Édition Presse Universitaire François-Rabelais, Tours, 381 p.

Medialdea, A., Thomsen, K.J., Murray, A.S., Benito, G., 2014. Reliability of equivalent-dose determination and age-models in the OSL dating of historical and modern palaeoflood sediments. Quaternary Geochronology 22, 11-24.

Murray, A.S, Thomsen, K.J., Masuda, N., Buylaert, J.P., Jain, M., 2012. Identifying well-bleached quartz using the different bleaching rates of quartz and feldspar luminescence signals. Radiation Measurements 47, 688-696.

Murray A.S. and Funder S., 2003. OSL dating of a Danish Eemian coastal marine deposit: a test of accuracy. Quaternary Science Reviews 22, 1177-1183.

Murray, A.S. and Wintle, A.G., 2003. The single aliquot regenerative dose protocol: potential for improvements in reliability. Radiation Measurements 37, 377-381.

Murray, A.S. and Olley, J.M., 2002. Precision and accuracy in the optically stimulated luminescence dating of sedimentary quartz: a status review. Geochronometria 21, 1-16.

Murray, A.S., Wintle A.G., 2000. Luminescence dating of quartz using an improved single-aliquot regenerative-dose protocol. Radiation Measurements 32, 57−73.

Murray, A.S., Marten, R., Johnston, A., Martin, P., 1987. Analysis for naturallyoccurring radionuclides at environmental concentrations by gamma spectrometry. Journal of Radioanalytical and Nuclear Chemistry 115, 263-288.

Olley, J.M., Pietsch, T., Roberts, R.G., 2004a. Optical dating of Holocene sediments from a variety of geomorphic settings using single grains of quartz. Geomorphology 60, 337-358.

Olley, J.M., Deckker, P.D., Roberts, R.G., Fifield, L.K., Yoshida, H., Hancock, G., 2004b. Optical dating of deep-sea sediments using single grains of quartz: a comparison with radiocarbon. Sedimentary Geology 169, 175-189.

Olley, J.M., Caitcheon, G.G., Roberts, R.G., 1999. The origin of dose distributions in fluvial sediments, and the prospect of dating single grains from fluvial deposits using optically stimulated luminescence. Radiation Measurements 30, 207-217.

Olley, J.M., Caitcheon, G.G., Murray, A.S., 1998. The distribution of apparent dose as determined by optically stimulated luminescence in small aliquots of fluvial quartz: implications for dating young sediments. Quaternary Geochronology 17, 1033-1040.

Olley, J.M., Roberts, R.G., Murray, A.S., 1997. Disequilibria in the uranium decay series in sedimentary deposits at Allen's cave, nullarbor plain, Australia: Implications for dose rate determinations. Radiation Measurements 27, 433–443.

Pietsch, T.J., 2009. Optically stimulated luminescence dating of young (<500 years old) sediments: testing estimates of burial dose. Quaternary Geochronology 4, 406–422.

Porat, N., Rosen, S.A., Boaretto, E., Avni, Y., 2006. Dating the Ramat Saharonim late Neolithic desert cult site. Journal of Archaeological Science 33, 1341-1355.

Prescott, J. R. & Hutton, J. T. 1994. Cosmic ray contributions to dose rates for luminescence and ESR dating: large depths and long-term variations. Radiation Measurements 23, 497–500.

Roberts, R.G., Galbraith, R.F., Olley, J.M., Yoshida, H., Laslett, G.M., 1999. Optical dating of single and multiple grains of quartz from Jinmium rock shelter, northern Australia: part II, results and implications. Archaeometry 41, 365-395.

Roberts, R.G., Bird, M., Olley, J., Galbraith, R., Lawson, E., Laslett, G., Yoshida, H., Jones, R., Fullagar, R., Jacobsen, G., Hua, Q., 1998. Optical and radiocarbon dating at Jinmium rock shelter in northern Australia. Nature 393, 358-362.

Sim, A.K., Thomsen, K.J., Murray, A.S., Jacobsen, G., Drysdale, R., Erskine, W., 2014. Dating recent floodplain sediments in the Hawkesbury-Nepean River system using single grain quartz OSL. Boreas 43, 1-21.

Spooner, N.A. and Allsop, A., 2000. The spatial variation of dose-rate from 90Sr/90Y beta sources for use in luminescence dating. Radiation Measurements 32, 49-56.

Tanner, A.B., 1980. Radon migration in the ground: a supplementary review. In Conference on natural radiation environment III, eds. Gesell, T.F.; Lowder, W.M., Department of Energy, Oak Ridge, TN (USA). CONF-780422-(VOL.1).

Thiel, C., Buylaert, J. P., Murray, A., Terhorst, B., Hofer, I., Tsukamoto, S., Frechen, M., 2011. Luminescence dating of the Stratzing loess profile (Austria) – Testing the potential of an elevated temperature post-IR IRSL protocol. Quaternary International 234, 23–31.

Thomsen, K.J., Murray, A.S., Buylaert, J. P., Jain, M. Submitted to Radiation Measurements. Application of the Fast Ratio, FR to single grain dose measurement obtained using laser stimulation.

Thomsen, K.J., Kook, M., Murray, A.S, Jain, M., Lapp, T., 2015. Single-grain results from an EMCCD-based imaging system. Radiation Measurements (in press).

Thomsen, K.J., Murray, A.S., Jain, M., 2012. The dose dependency of the over-dispersion of quartz OSL single grain dose distributions. Radiation Measurements 47, 732-739.

Thomsen, K. J., Murray, A. S., Bøtter-Jensen, L., Kinahan, J., 2007. Determination of burial dose in incompletely bleached fluvial samples using single grains of quartz. Radiation Measurements 42, 370–379.

Thomsen, K.J., Murray, A.S., Bøtter-Jensen, L., 2005. Sources of variability in OSL dose measurements using single grains of quartz**.** Radiation Measurements 39, 1, 47-61.

Wintle, A.G., Murray, A.S., 2006. A review of quartz optically stimulated luminescence characteristics and their relevance in single-aliquot regeneration dating protocols. Radiation Measurements 41, 369 – 391.

Yoshida, H., Roberts, R.G., Olley, J.M., Laslett, G.M., Galbraith, R.F., 2000. Extending the age range of optical dating using single ‘supergrains’ of quartz. Radiation Measurements 32, 439–446.

Zhao, Q.Y., Thomsen, K.J., Murray, A.S., Wei, M.J., Pan, B.L., Zhou, R., Zhao, X.H., Chen, H.Y., 2015. Quartz single-grain OSL dating of debris flow deposits from Miyun, north east Beijing in China. Quaternary Geochronology (in press).

Figure captions

Figure 1: Composite cross-sections of the Bordes-Fittes rockshelter (adapted from Aubry et al., 2012). Each cross-section is located (as a thick black line) on the plan view shown in Figure S1. The legend identifies the various units described by Aubry et al. (2012; 2014). Locations of OSL samples are given as filled red squares and ^{14}C samples as filled blue circles; all these samples are individually labelled for identification.

Figure 2: Effect of single-grain (SG) rejection criteria on the number of accepted grains (n), the Central Age Model (CAM) dose and the relative over-dispersion (OD) normalised to the values obtained after application of the first rejection criterion, e.g. the response to the first test dose must be known to better than 20% ("s_{TN}<20 %"). "Recyc." is the recycling rejection criteria (the recycling ratio must be consistent with unity within two standard deviations), "IR depl." is the IR depletion rejection criteria (the IR depletion ratio must be consistent with unity within 2 standard deviations) and "Recup" is the recuperation rejection criteria (the recuperation must be smaller than or consistent with 1 Gy). The squares represent results from nine individual samples - both natural and laboratory gamma irradiated (74.8 Gy) dose distributions for 092201, -02, -03 and -04 and a further laboratory beta irradiated (65 Gy) dose distribution for sample 092204. The circles, diamonds and triangles represent average values for the nine samples.

Figure 3: Comparison of the results for the four OSL samples obtained using OSL multi-grain quartz and feldspar (Table 3) and quartz OSL single-grains using the standard approaches and rejection criteria (section 3.2) with the independent age control provided by ^{14}C (see Table 1). All ages have been converted into predicted quartz doses for ease of comparison. The doses based on ^{14}C ages most relevant to the OSL samples investigated here are shown as dark grey bands and are derived from the outermost 95% confidence limit of the relevant ^{14}C age multiplied by the appropriate quartz dose rate without including dose rate uncertainty. The light grey bands then include the additional uncertainty from the dose rate term. The ^{14}C age control for OSL samples 092201 and -02 are based on the arithmetic average of five ^{14}C ages (i.e. OxA-22316, -22342, -226470, -226471 and -226472) for unit GFU D2 (see text for details). The ^{14}C age control for OSL samples 092203 and -04 are based on two ^{14}C ages (i.e. OxA-22315 and Beta-234193). For OSL sample 092204 the quartz dose must lie between the dose predicted from these ^{14}C ages (indicated by black arrows in the figure). For OSL sample 092203, ^{14}C sample Beta-234193 is expected to provide only an upper limit; in contrast the ^{14}C sample OxA-22315 was sample immediately adjacent to the OSL sample and it is considered likely that the quartz dose for this sample should lie close to that predicted from the ^{14}C age. Although there is an additional systematic uncertainty in deriving quartz doses from K-feldspar ages arising from the internal dose rate, this is small and has been neglected here. All single-grain burial dose estimates have been calculated using an additional uncertainty of 15% (see text for details).

Figure 4: Multi-grain (a-c) and single-grain (d-f) quartz luminescence characteristics. a), d) Normalised dose response curves for sample 092201. The open symbols indicate recycling points. The numbers given next to the individual dose response curves in d) indicate the respective D_0 values. The natural sensitivity corrected signals are indicated at x=0. The inset in a) shows the normalised OSL signals measured for the natural and regenerated (80 Gy) signals. The normalised OSL signal measured from calibration quartz is shown for comparison. The signal (grey shaded) and background (hatched) summation intervals are also indicated. The inset in d) shows normalised and background subtracted single-grain stimulation curves. b), e) Frequency histograms of the D_0 values for individual DRCs for accepted natural aliquots for sample 092201. In e) only D_0 values known to better than 50% are shown. c), f) Natural average dose as a function of preheat temperature for samples 092202 (squares) and 092203 (circles). For multi-grain measurements each point is the average of the doses determined for 8 individual aliquots. For single-grain measurements each point is the CAM average of at least 33 individual single-grain accepted dose estimates. The thermal treatment (cutheat) prior to the test dose cycle was kept constant at 220 °C.

Figure 5: Natural single-grain quartz dose distributions. The assigned uncertainties are based on counting statistics, curve fitting errors and an intrinsic source of uncertainty of 15% per dose measurement. In the scatter plots of intrinsic brightness (natural test dose response) against estimated dose, and in the radial plots, the open symbols represent those dose estimates that are rejected based on the two additional/alternative rejection criteria based on D_0 and F_R values (see text for details). In the frequency histograms all doses have been included irrespective of the precision with which they are known. The red solid lines shown in the scatter plots, as well as in the radial plots, indicate the components identified by the FMM (see Table 6). Note that the appearance of radial plots is highly dependent on the uncertainty assigned to individual dose values (see Figure S5).

Figure 6: Doses estimated by the Finite Mixture Model (FMM) as a function of additional uncertainty for all four natural samples. This uncertainty is additional to the uncertainty estimated based solely on counting statistics and dose response curve fitting errors. Squares represent the burial dose derived from the most prominent grain population (1^{st}), i.e. the component containing the largest fraction of the individual dose estimates. Circles represent the burial dose estimated by the second largest (2^{nd}) component. The proportions of grains given by the FMM are given adjacent to some of the individual dose estimates. The horizontal dashed lines are the single-grain CAM doses and the shaded horizontal bars represent the expected dose (based on multi-grain quartz results).

Figure 7: Distribution of D_0 values for accepted single-grain dose estimates for sample 092201 given a gamma dose of 107.4 Gy. The D_0 values for 10 grains (out of 119) have been omitted, because the dose response curves for these grains are close to linear and subsequently give very large D_0 values with very poor precisions. The top part of the graph shows the corresponding cumulative frequency distribution. The vertical line indicates a D_0 value of half of the given dose (e.g. 53.7 Gy). The inset shows the effect of omitting dose estimates based on their D_0 value on the dose recovery ratio for the average (arithmetic) dose (squares) and the CAM dose (circles). The dashed horizontal lines indicate the plateau value (i.e. the mean value of the dose recovery ratios for D_0>100-200 Gy) for the two dose estimation methods, respectively.

Figure 8: The effect of omitting single-grain dose estimates based on their D_0 values for the four natural samples. The average (arithmetic) dose is shown by the filled squares and the CAM dose by the open circles. The grey horizontal bands indicate the multi-grain quartz doses for the individual samples and the dashed lines indicate when the dose is equal to the smallest D_0 value.

Figure 9: The effect of various rejection criteria on the single-grain arithmetic averages (filled squares) and the CAM doses (open circles). "Std." refers to the standard rejection criteria outlined in section 3.2. "D_0>x" is the additional rejection criterion detailed in section 8.4.2, i.e. the D_0 value of the DRC for a given grain must be larger than x for the grain to be accepted into the dose distribution. "F_R>4" is the additional rejection criteria given in section 8.4.3, i.e. the decay rate quantified by the fast ratio, F_R, for any accepted grain must be larger than 4. "D_0 & F_R" combines the three above mentioned rejection criteria (see section 8.4.4). The horizontal grey bars represent the measured multi-grain quartz doses whereas the horizontal hatched bars represent the quartz dose predicted by the $pIRIR_{290}$ feldspar measurements (as given in Aubry et al., 2012).

Figure 1

Click here to download high resolution image

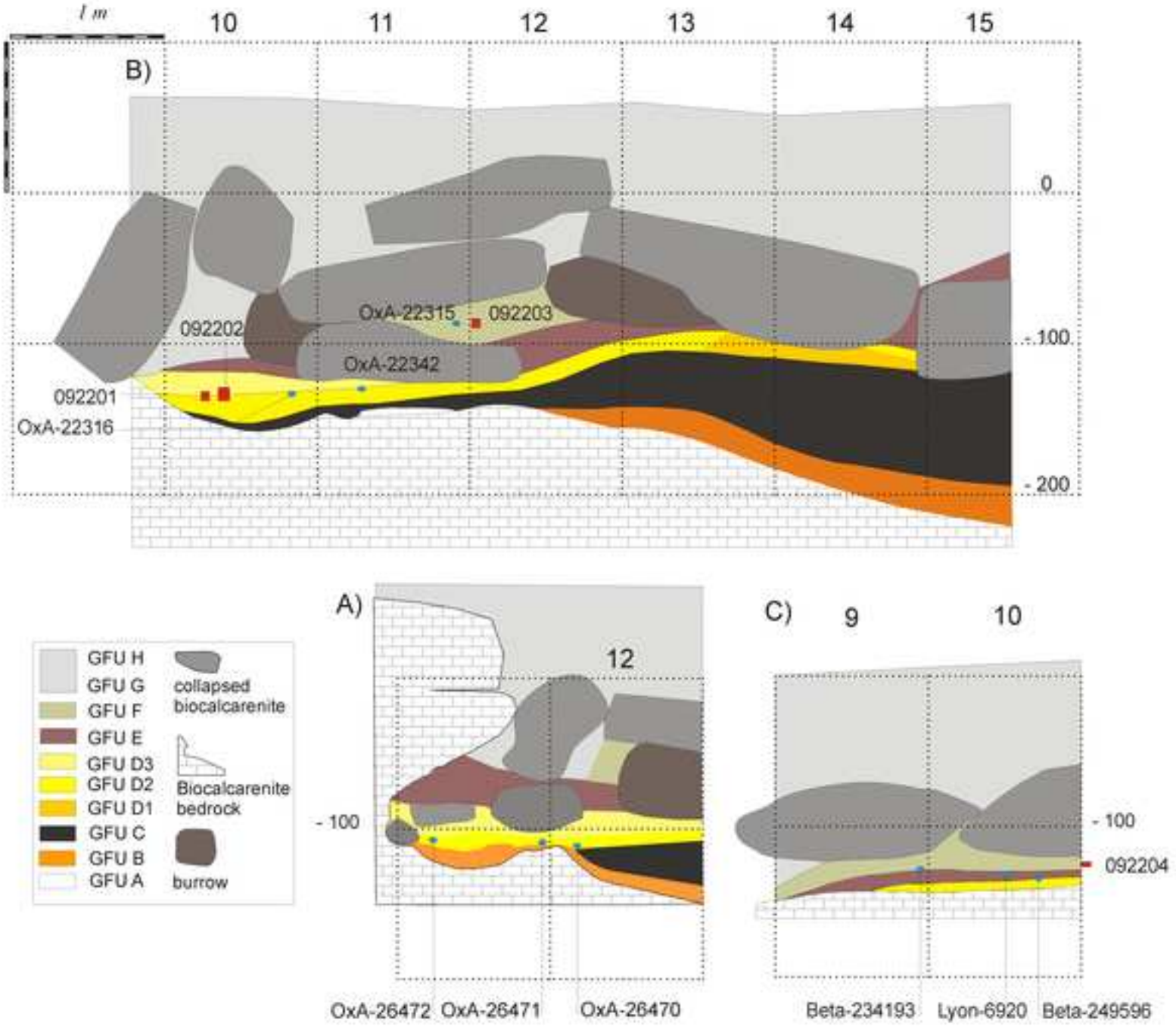


Figure 2

Click here to download high resolution image

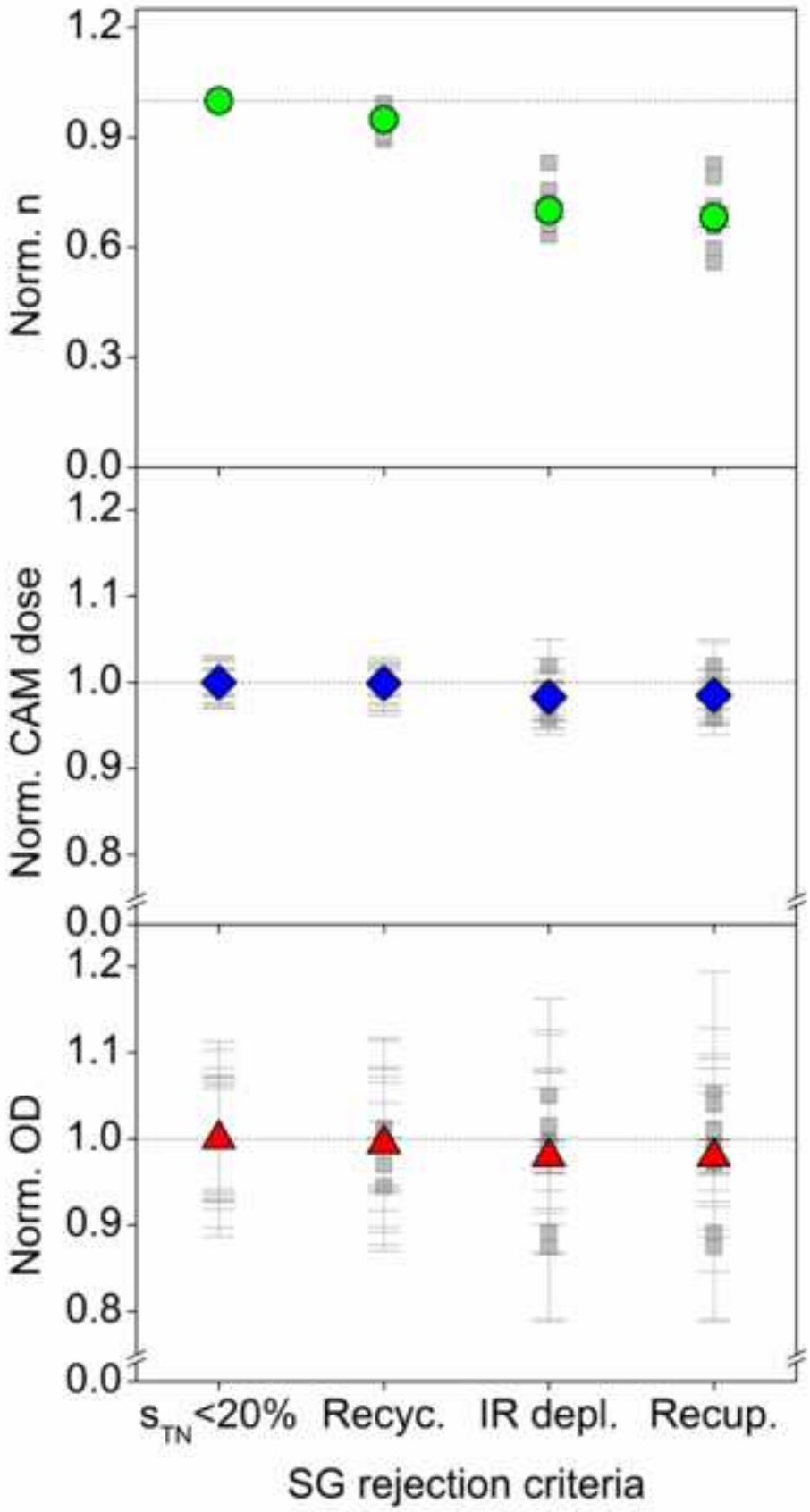


Figure 3

Click here to download high resolution image

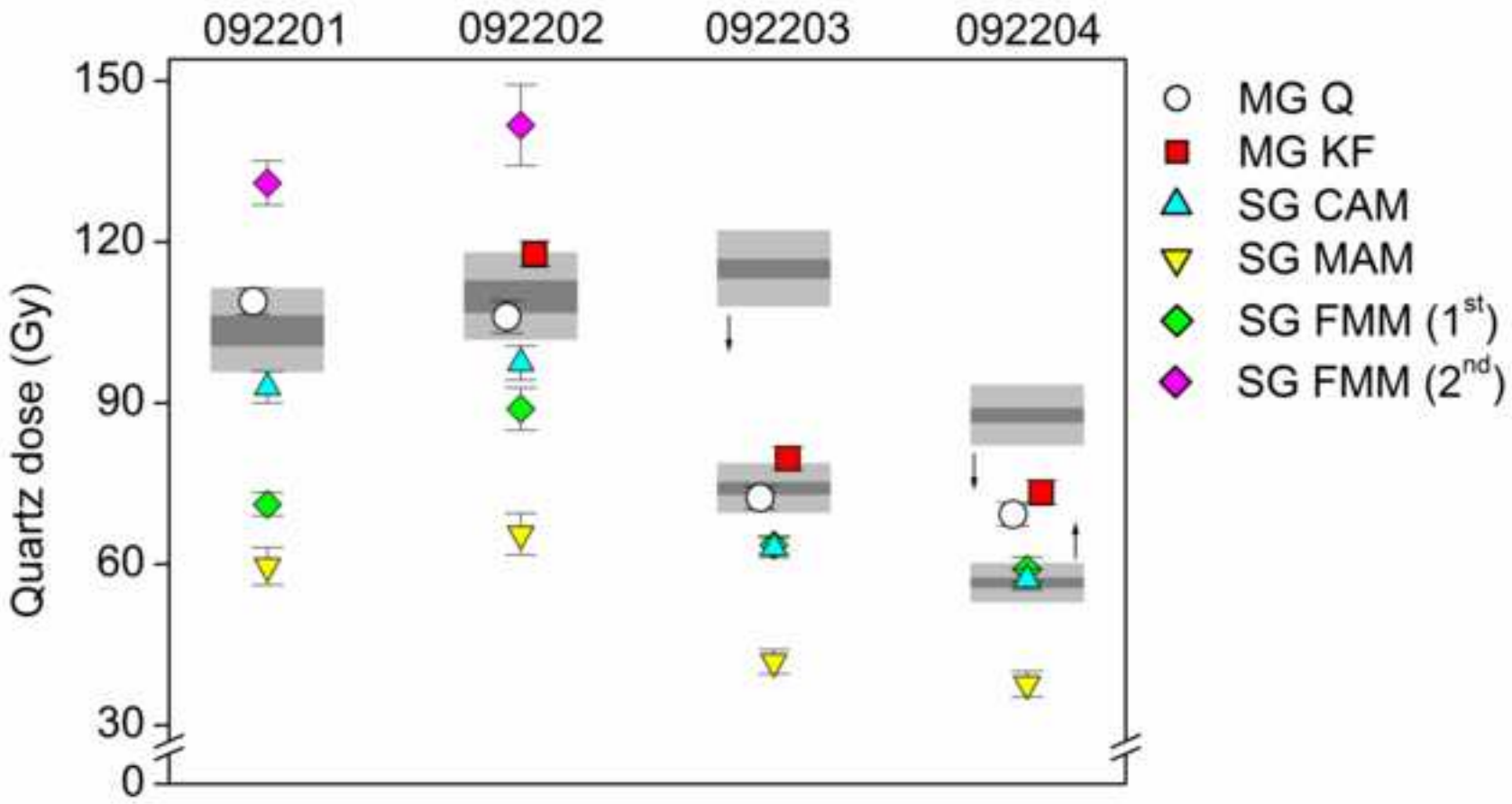


Figure 4

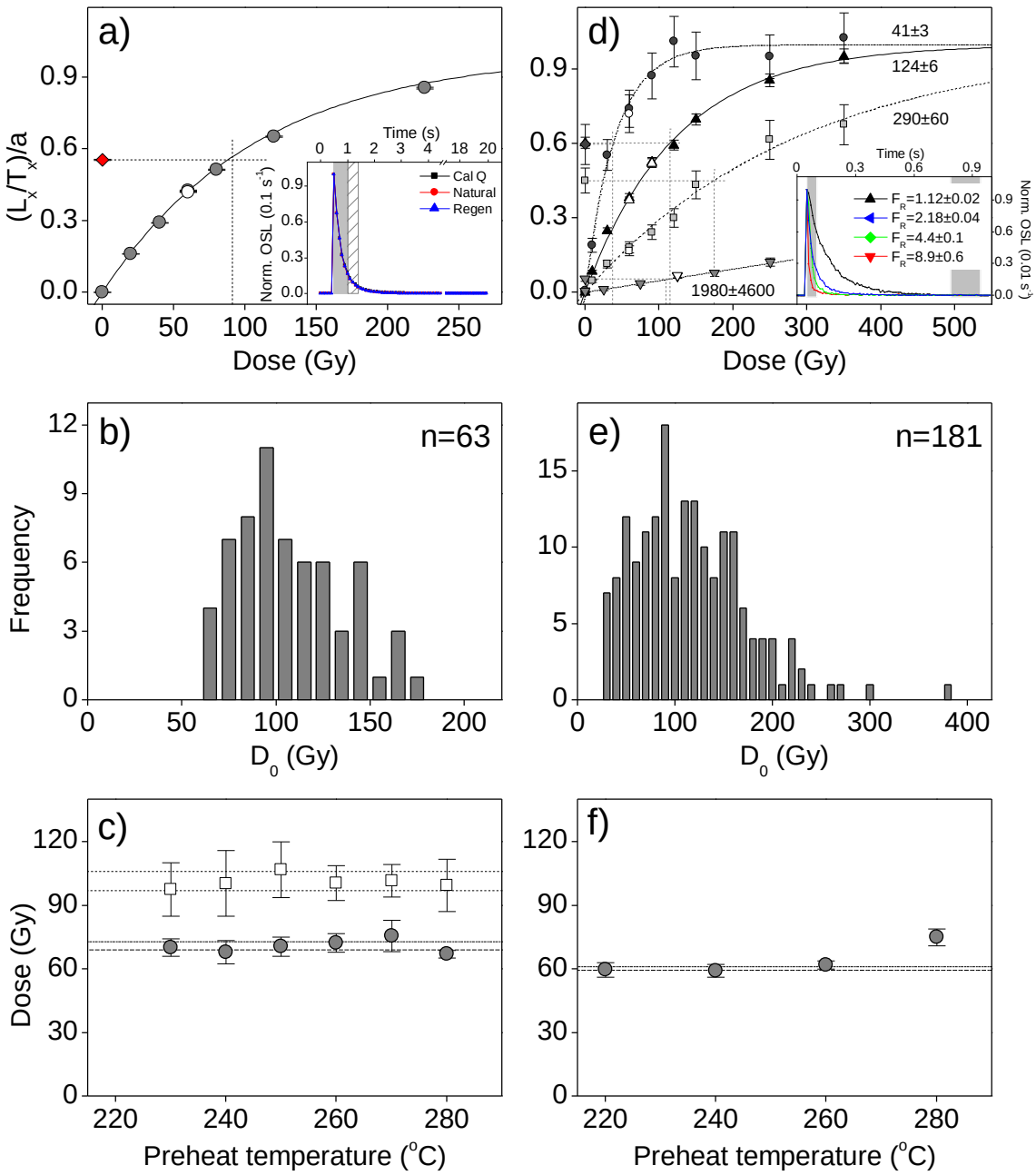

a)
$(L_x/T_x)/a$
Dose (Gy)
Time (s)
Norm. OSL (0.1 s^{-1})
Cal Q
Natural
Regen
d)
41±3
124±6
290±60
1980±4600
Time (s)
F_R=1.12±0.02
F_R=2.18±0.04
F_R=4.4±0.1
F_R=8.9±0.6
Norm. OSL (0.01 s^{-1})
Dose (Gy)
b)
n=63
Frequency
D_0 (Gy)
e)
n=181
D_0 (Gy)
c)
Dose (Gy)
Preheat temperature (°C)
f)
Preheat temperature (°C)

Figure 5

Figure 5

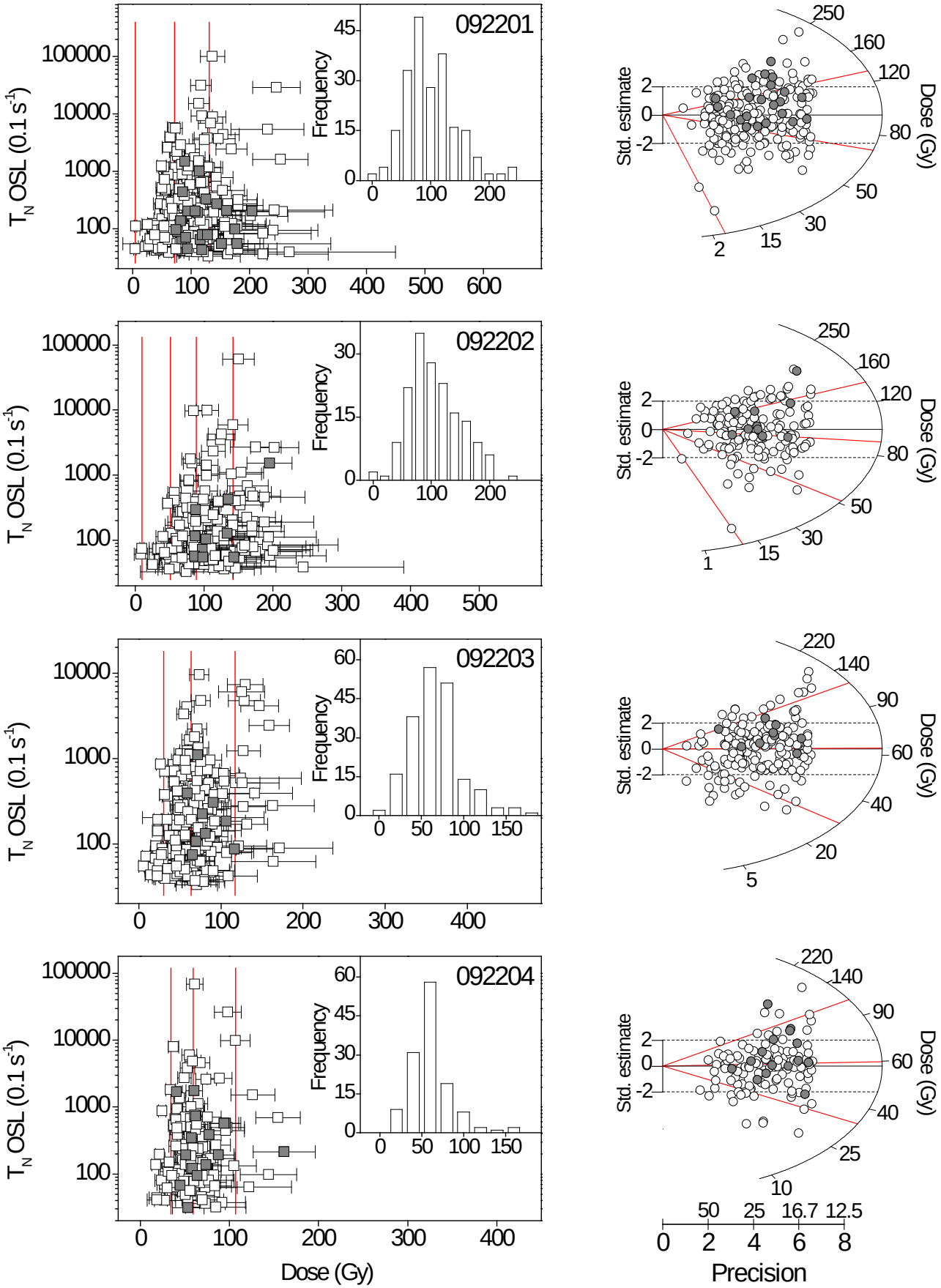


Figure 6

Click here to download high resolution image

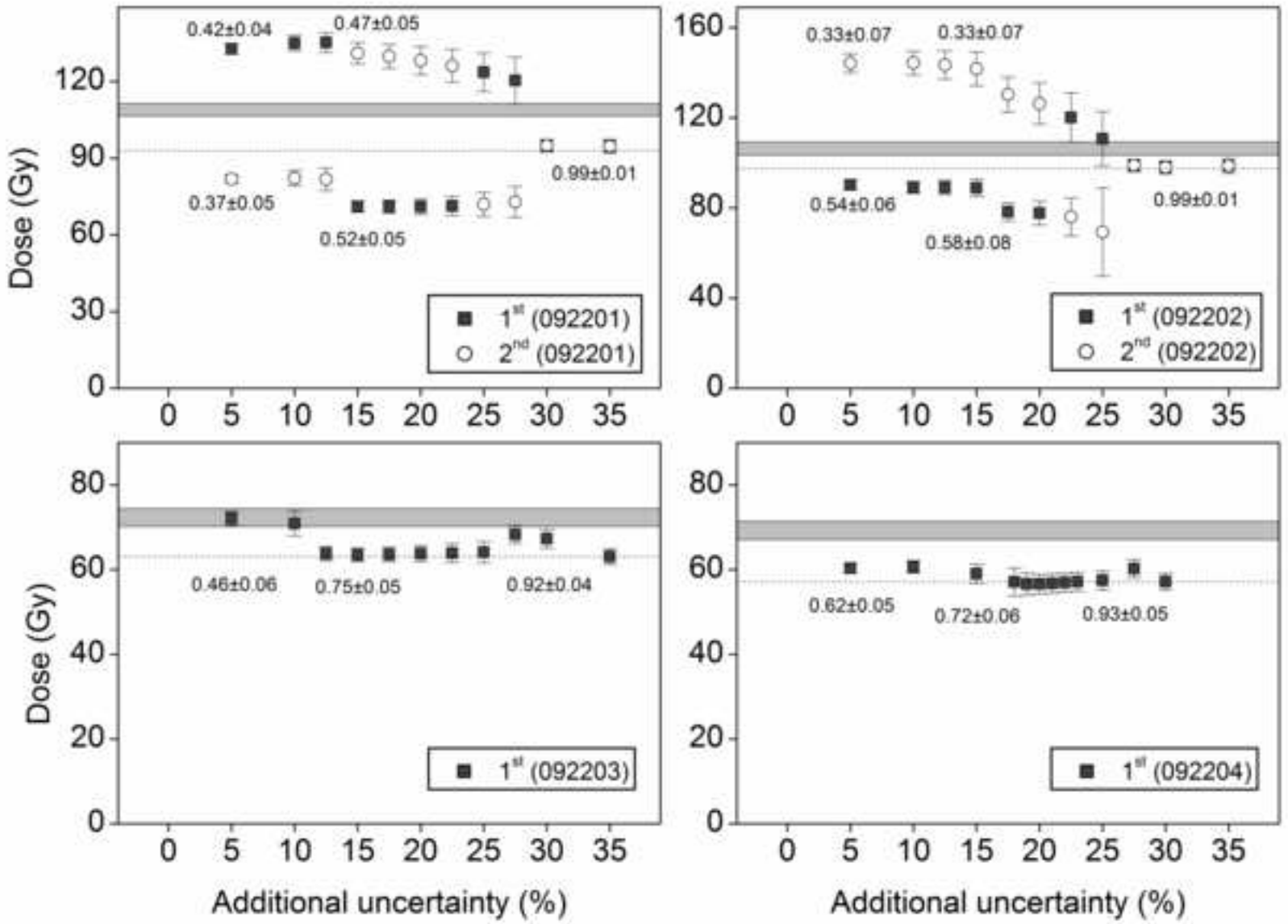


Figure 7

Click here to download high resolution image

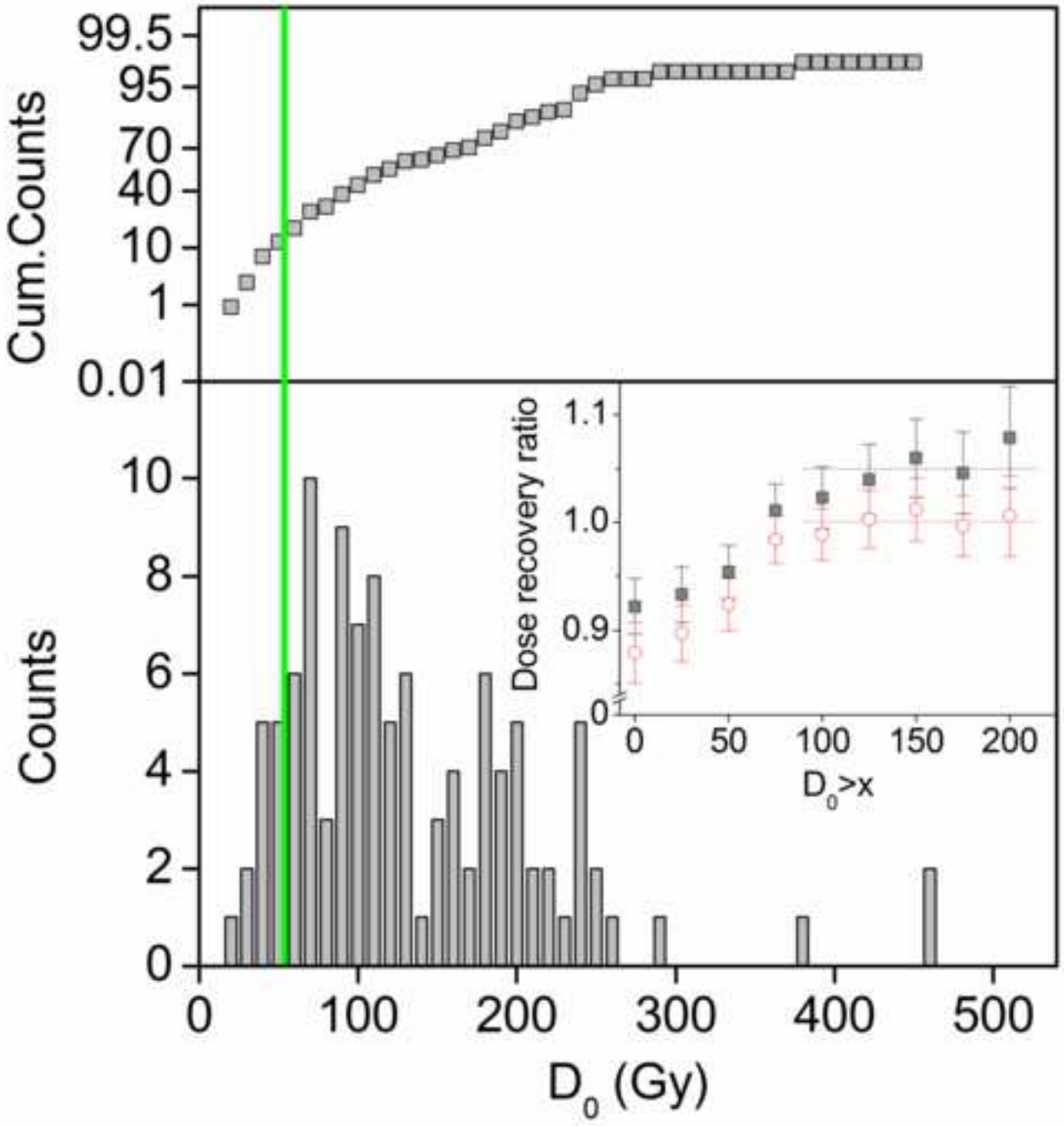


Figure 8

Click here to download high resolution image

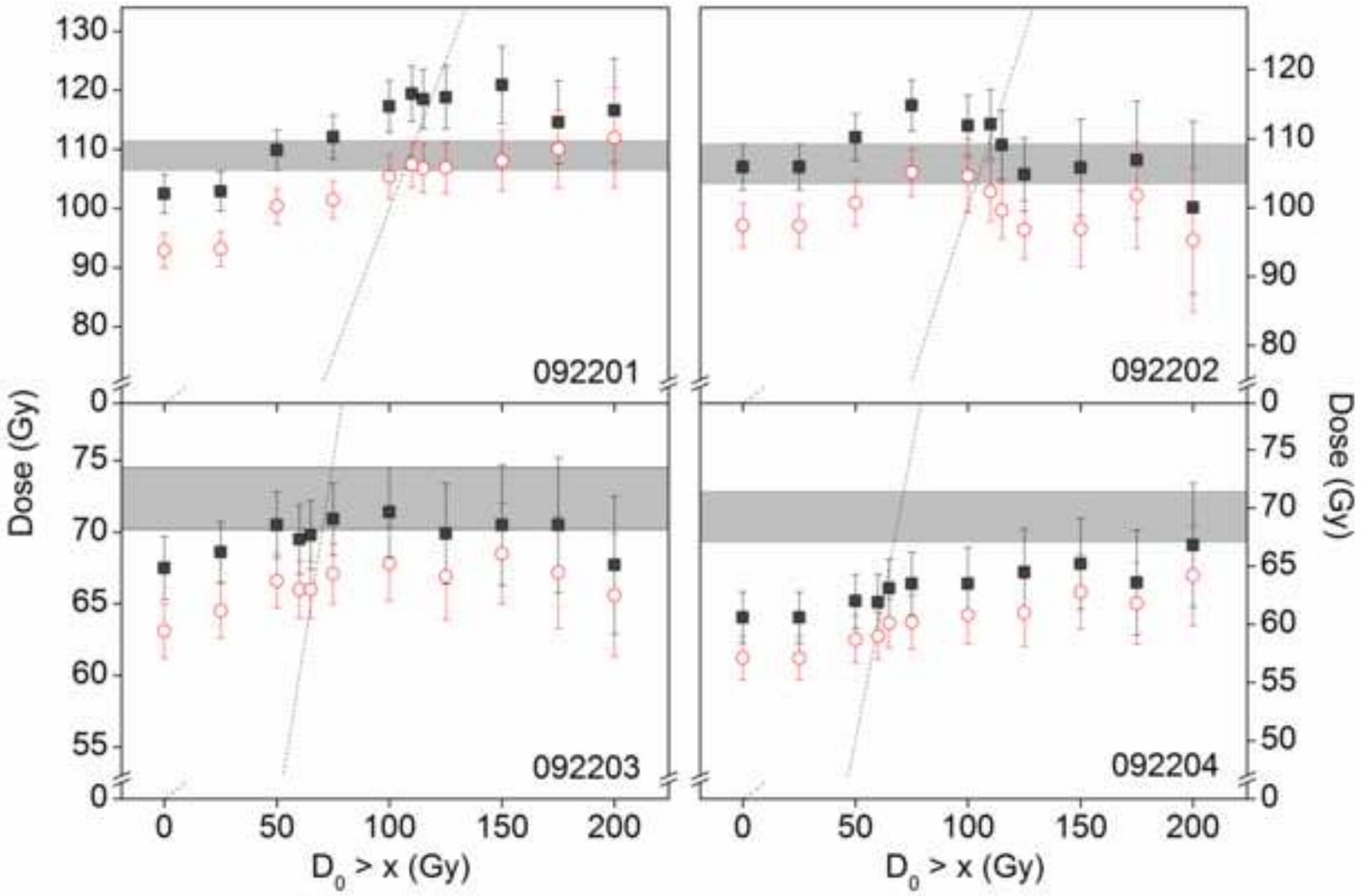
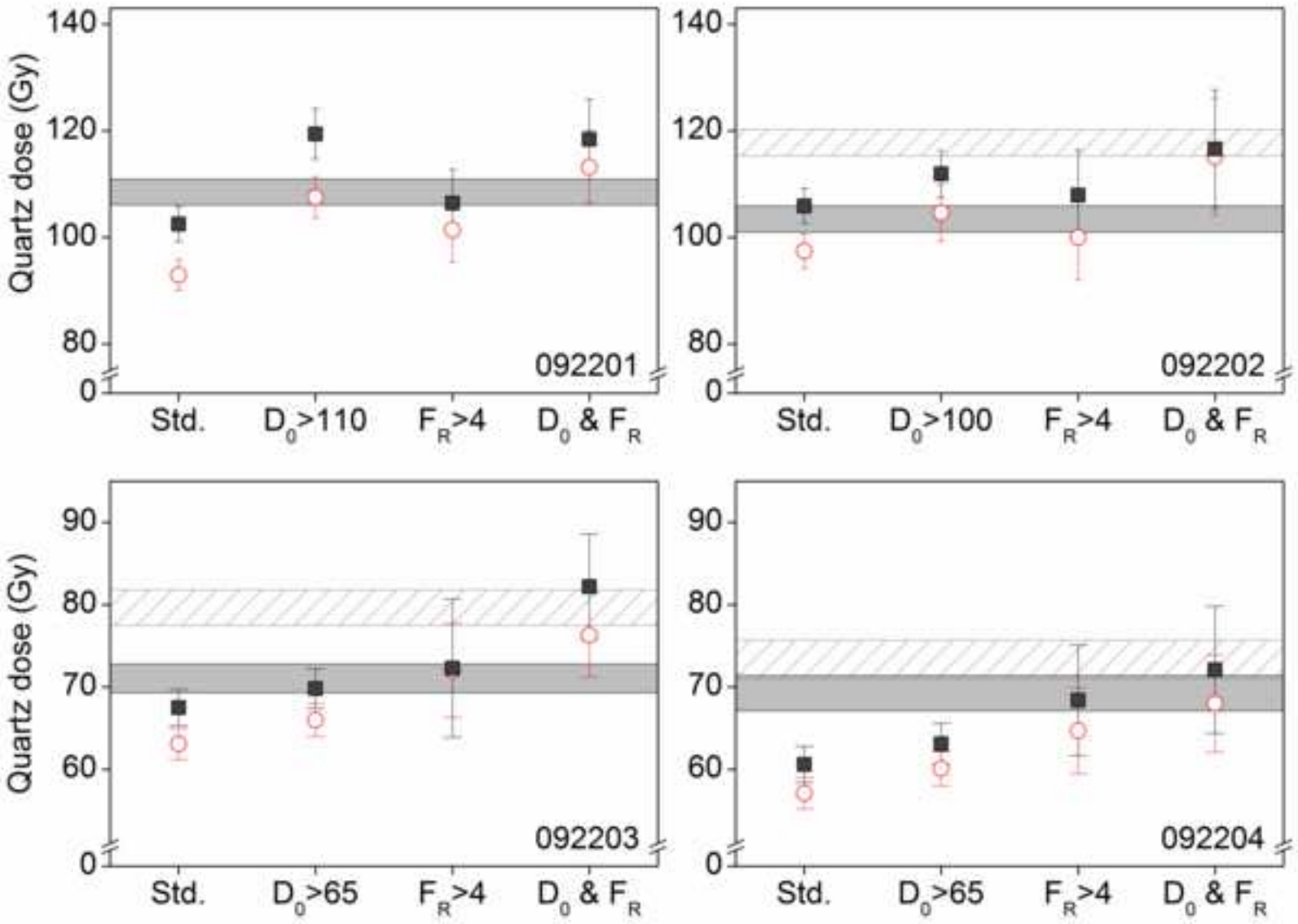


Figure 9

Click here to download high resolution image



Table

Table 1. ^{14}C AMS results

Lab code	Unit	Material	Pre-treatment[a,b]	^{14}C uncalibrated age (yr BP)	^{14}C calibrated age (cal yr)
OxA-22315	GFU F	Antler	U.F.	19020 ± 110	23,322-22,615
Beta-234193	GFU F	Bone	Coll.	31640 ± 230	36,137-35,059
Lyon-6920	GFU E	Bone	U.F.	34520 ± 850	41,155-36,985
Beta-249596	GFU E	Bone	Coll.	35770 ± 380	41,339-39,631
Beta-249595	GFU D2	Bone	Coll.	32110 ± 280	36,684-35,420
OxA-22342	GFU D2	Tooth	U.F.	37400 ± 800	43,106-40,506
OxA-22316	GFU D2	Bone	U.F.	41200 ± 1300	47,864-42,799
OxA-26470	GFU D2	Bone	U.F.	39100 ± 1000	45,020-41,810
OxA-26471	GFU D2	Bone	U.F.	41000 ± 1300	47,654-42670
OxA-26472	GFU D2	Bone	U.F.	40600 ± 1200	46,838-42,515
OxA-29527	GFU D1	Bone	U.F.	41900 ± 1500	49,036-43,161

These data (95.4% probability) are taken from Aubry et al. (2012;2014). Note the calibrated ages are not identical to published values because some required recalibration (OxCal v.4.2 calibration program; Bronk Ramsey, 2013) and all are referrred to 2009 for comparison with OSL ages.

[a] U.F. is ultrafiltration

[b] Coll. is Collagen extraction with alkali

Table 2. Summary of radionuclide concentrations and derived quartz dose rates

Sample	Depth	w.c.	^{238}U	^{226}Ra	^{232}Th	^{40}K	Dry β dose rate	Dry γ dose rate	Total dose rate
	cm	%	$Bq.kg^{-1}$	$Bq.kg^{-1}$	$Bq.kg^{-1}$	$Bq.kg^{-1}$	$Gy.ka^{-1}$	$Gy.ka^{-1}$	$Gy.ka^{-1}$
092201	200	1	14 ± 6	20.3 ± 0.5	23.4 ± 0.6	497 ± 11	1.59 ± 0.04	0.81 ± 0.02	2.36 ± 0.11
092202	200	3	15 ± 6	23.9 ± 0.4	27.9 ± 0.5	517 ± 10	1.69 ± 0.04	0.90 ± 0.02	2.51 ± 0.12
092203	170	4	26 ± 4	36.5 ± 0.4	42.2 ± 0.5	658 ± 11	2.26 ± 0.04	1.27 ± 0.03	3.23 ± 0.15
092204	150	1	26 ± 6	22.6 ± 0.5	25.5 ± 0.6	557 ± 12	1.82 ± 0.04	0.90 ± 0.02	2.47 ± 0.12
TA2265R			-2 ± 14	5.8 ± 1	7.2 ± 1.0	60 ± 12		0.18 ± 0.02	
TA2266R			17 ± 17	10.3 ± 1.1	4.9 ± 1.0	30 ± 14		0.17 ± 0.02	
TA2267R			-2 ± 13	6.2 ± 0.9	8.9 ± 0.9	48 ± 10		0.196 ± 0.013	
TA2268R			19 ± 11	15.9 ± 0.8	6.6 ± 0.7	66 ± 9		0.268 ± 0.012	

Note: Sediment samples are termed 092201-04. Rock fragment samples are termed TA2265R-68R. Water content (w.c.) taken as observed field water content, expressed with respect to sample dry weight. In dose rate calculations we have associated a ±4% uncertainty on these water contents. Total dose rate includes cosmic ray component and the effects of water content. Uranium series equilibrium assumed when calculating dry gamma dose rates for the four rock samples. Assumptions for the sediment samples are discussed in text.

Table 3. OSL multi-grain results

Lab code	Unit	Mineral	Dose (Gy)	n	Dose rate (Gy/ka)	Age (ka)
092201	D2	Quartz	109 ± 2	63	2.36 ± 0.11	46.1 ± 2.4
092202	D2	Quartz	106 ± 3	50	2.51 ± 0.12	42.4 ± 2.3
	D2	Feldspar	156 ± 3	8	3.32 ± 0.12	46.9 ± 1.9
092203	F	Quartz	72 ± 2	74	3.23 ± 0.15	22.4 ± 1.2
	F	Feldspar	100 ± 2	9	4.04 ± 0.15	24.6 ± 1.1
092204	F	Quartz	69 ± 2	66	2.47 ± 0.12	28.1 ± 1.6
	F	Feldspar	97 ± 3	9	3.28 ± 0.13	29.7 ± 1.4

Unit is the stratigraphic unit. The feldspar and quartz data have been taken from Aubry et al. (2012; 2014).

Table 4. Summary of single-grain dose recovery results

Bleaching	Irradiation	Given dose (Gy)	Sample	N[a]	Standard rejection criteria					$D_0 > D_{given}$ criterion			
					n[b]	n_{SAT}[c]	D.R.[d]	OD (%)[e]	Model[f]	n[b]	n_{SAT}[c]	D.R.[d]	OD (%)[e]
SOL2	^{60}Co	107.4	092201	1000	119	23%	0.88 ± 0.03	29 ± 3	CAM	70	1%	0.99 ± 0.02	14 ± 2
			092202	1900	192	38%	1.04 ± 0.03	22 ± 2	MAM-4	118	2%	1.07 ± 0.03	20 ± 2
		74.8	092201	1800	150	9%	0.91 ± 0.02	18 ± 2	MAM-3	109	1%	0.95 ± 0.02	17 ± 2
			092202	2400	186	11%	1.00 ± 0.02	14 ± 2	MAM-3	132	1%	1.04 ± 0.02	12 ± 2
			092203	1800	144	10%	0.88 ± 0.02	20 ± 1	MAM-4	90	0%	0.95 ± 0.02	14 ± 2
			092204	1200	79	4%	0.91 ± 0.03	19 ± 2	MAM-3	49	0%	0.98 ± 0.03	13 ± 3
		28.1	092201	1800	178	8%	1.02 ± 0.02	20 ± 2	MAM-4	168	2%	1.02 ± 0.02	20 ± 2
			092202	1700	162	3%	1.04 ± 0.02	18 ± 1	MAM-3	157	1%	1.05 ± 0.02	19 ± 2
			092203	1100	97	2%	1.00 ± 0.02	12 ± 2	MAM-3	88	0%	1.01 ± 0.02	12 ± 2
			092204	1200	92	2%	1.00 ± 0.02	18 ± 2	L5%	90	0%	1.01 ± 0.02	18 ± 2
		9.5	092201	1800	146	0%	0.97 ± 0.02	17 ± 2	MAM-3	146	0%	0.97 ± 0.02	17 ± 2
		4.9	092201	1700	92	0%	1.03 ± 0.02	15 ± 2	MAM-3	92	0%	1.03 ± 0.02	15 ± 2
SOL2	$^{90}Sr/^{90}Y$	120	092201	1200	141	26%	0.90 ± 0.02	25 ± 2	MAM-4	64	2%	1.01 ± 0.03	16 ± 3
		110	092201	1200	138	31%	1.01 ± 0.03	24 ± 2	CAM	67	1%	1.10 ± 0.03	14 ± 3
		100	092201	1200	137	14%	0.94 ± 0.02	22 ± 2	CAM	68	1%	1.06 ± 0.02	13 ± 2
		80	092201	1200	138	20%	1.03 ± 0.03	24 ± 2	CAM	85	1%	1.10 ± 0.02	14 ± 2
		65	092201	1800	103	9%	1.01 ± 0.02	18 ± 2	CAM	79	1%	1.05 ± 0.02	15 ± 2
			092204	1200	71	8%	1.02 ± 0.02	12 ± 2	MAM-3	46	0%	1.06 ± 0.02	12 ± 2
		28	092201	900	115	0%	0.98 ± 0.02	17 ± 2	MAM-3	93	0%	1.00 ± 0.02	15 ± 2
		10	092201	1000	111	0%	1.07 ± 0.02	16 ± 2	MAM-3	111	0%	1.07 ± 0.02	16 ± 2
Blue LEDs	$^{90}Sr/^{90}Y$	110	092202	1200	105	21%	0.97 ± 0.03	25 ± 3	CAM	51	6%	1.07 ± 0.03	10 ± 2
			092204	1200	83	19%	0.96 ± 0.02	10 ± 2	MAM-3	45	0%	0.97 ± 0.02	5 ± 2
		85	092201	1200	135	12%	1.03 ± 0.02	19 ± 2	MAM-3	101	1%	1.06 ± 0.02	14 ± 2
		65	092201	1100	165	8%	0.96 ± 0.02	21 ± 2	MAM-3	122	0%	1.03 ± 0.02	15 ± 2
		28	092202	1200	115	2%	1.02 ± 0.02	19 ± 2	L5%	113	0%	1.02 ± 0.02	19 ± 2
			092204	1200	106	0%	1.04 ± 0.01	7 ± 1	MAM-4	105	0%	1.04 ± 0.01	7 ± 1
		10	092201	1200	93	0%	1.06 ± 0.02	12 ± 2	CAM	90	0%	1.06 ± 0.02	12 ± 2
		4.8	092201	900	37	0%	1.01 ± 0.03	10 ± 3	MAM-3	37	0%	1.01 ± 0.03	10 ± 3

[a] Number of measured grains

[b] number of accepted dose estimates

[c] number of grains in saturation

[d] dose recovery ratio calculated using the CAM dose

[e] the relative over-dispersion when the assigned uncertainites are solely based on counting statistics and curve fitting errors.

[f] the dose estimation model predicted by the decision tree proposed by Bailey and Arnold (2006).

Table 5. Summary of quartz multi-grain (MG) and single-grain (SG) results.

Sample	MG					SG									
	Dose (Gy)		n[a]	σ (%)[b]	OD (%)	N[c]	Rejection criteria[d]	n[a]	n/N	n in SAT[e]		Dose (Gy)		OD (%)[f]		Measured/Expected dose[g]	
	Average	CAM										Average	CAM	c.s.	c.s. +15%	Average	CAM
092201	109 ± 2	107 ± 2	63	18	17 ± 2	2300	Std.	218	9%	55	(20%)	103 ± 3	93 ± 3	40 ± 3	38 ± 3	0.94 ± 0.04	0.85 ± 0.03
							s_{Tn}<20%	325	14%	66	(17%)	101 ± 3	93 ± 2	38 ± 2	35 ± 2	0.93 ± 0.03	0.85 ± 0.03
							Std. & D_0>110	108	5%	1	(1%)	119 ± 5	107 ± 4	32 ± 3	29 ± 3	1.10 ± 0.05	0.99 ± 0.04
							s_{Tn}<20% & D_0>110	176	8%	2	(1%)	112 ± 4	103 ± 3	33 ± 2	29 ± 2	1.03 ± 0.04	0.94 ± 0.03
							Std. & F_R>4	36	2%	6	(14%)	106 ± 6	101 ± 6	30 ± 5	26 ± 5	0.98 ± 0.06	0.93 ± 0.06
							s_{Tn}<20% & F_R>4	53	2%	7	(12%)	103 ± 5	98 ± 5	29 ± 4	25 ± 3	0.94 ± 0.05	0.90 ± 0.05
							Std. & D_0>110 & F_R>4	23	1%	0	(0%)	118 ± 8	113 ± 7	23 ± 5	18 ± 4	1.09 ± 0.07	1.04 ± 0.07
							s_{Tn}<20% & D_0>110 & F_R>4	33	1%	0	(0%)	111 ± 6	105 ± 6	29 ± 5	25 ± 4	1.02 ± 0.06	0.96 ± 0.06
092202	106 ± 3	105 ± 3	50	19	17 ± 2	2400	Std.	167	7%	52	(24%)	105 ± 3	97 ± 3	36 ± 3	32 ± 2	0.99 ± 0.04	0.91 ± 0.04
							s_{Tn}<20%	205	9%	69	(25%)	108 ± 3	99 ± 3	37 ± 2	34 ± 2	1.01 ± 0.04	0.93 ± 0.04
							Std. & D_0>100	92	4%	7	(7%)	111 ± 5	103 ± 4	31 ± 3	27 ± 3	1.04 ± 0.05	0.96 ± 0.04
							s_{Tn}<20% & D_0>100	109	5%	7	(6%)	113 ± 4	99 ± 3	33 ± 3	30 ± 3	1.06 ± 0.05	0.93 ± 0.04
							Std. & F_R>4	21	1%	3	(13%)	108 ± 9	99 ± 8	30 ± 6	26 ± 5	1.01 ± 0.08	0.93 ± 0.08
							s_{Tn}<20% & F_R>4	23	1%	3	(12%)	104 ± 9	101 ± 8	30 ± 6	26 ± 5	0.98 ± 0.09	0.95 ± 0.08
							Std. & D_0>100 & F_R>4	11	0%	0	(0%)	106 ± 14	113 ± 11	25 ± 7	20 ± 6	1.00 ± 0.14	1.07 ± 0.11
							s_{Tn}<20% & D_0>100 & F_R>4	11	0%	0	(0%)	106 ± 14	113 ± 11	25 ± 7	20 ± 6	1.00 ± 0.14	1.07 ± 0.11
092203	72 ± 2	71 ± 2	74	25	22 ± 2	1800	Std.	195	11%	23	(11%)	67 ± 2	63 ± 2	36 ± 3	33 ± 2	0.93 ± 0.04	0.87 ± 0.04
							s_{Tn}<20%	257	14%	30	(10%)	66 ± 2	62 ± 2	36 ± 2	33 ± 2	0.91 ± 0.04	0.85 ± 0.03
							Std. & D_0>65	143	8%	1	(1%)	70 ± 2	66 ± 2	31 ± 2	28 ± 2	0.96 ± 0.04	0.91 ± 0.04
							s_{Tn}<20% & D_0>65	194	11%	3	(2%)	67 ± 2	64 ± 2	32 ± 2	28 ± 2	0.93 ± 0.04	0.88 ± 0.03
							Std. & F_R>4	11	1%	3	(21%)	72 ± 8	72 ± 6	19 ± 10	11 ± 6	1.00 ± 0.12	1.00 ± 0.08
							s_{Tn}<20% & F_R>4	14	1%	3	(18%)	73 ± 8	71 ± 5	20 ± 9	13 ± 6	1.01 ± 0.11	0.98 ± 0.08
							Std. & D_0>65 & F_R>4	9	1%	1	(10%)	82 ± 6	76 ± 5	13 ± 6	—	1.14 ± 0.09	1.05 ± 0.08
							s_{Tn}<20% & D_0>65 & F_R>4	12	1%	1	(8%)	81 ± 6	74 ± 5	15 ± 6	0 ± 6	1.12 ± 0.10	1.03 ± 0.07
092204	69 ± 2	67 ± 2	66	26	26 ± 2	1600	Std.	130	8%	16	(11%)	61 ± 2	57 ± 2	35 ± 3	31 ± 2	0.87 ± 0.04	0.83 ± 0.04
							s_{Tn}<20%	192	12%	23	(11%)	61 ± 2	57 ± 2	35 ± 2	32 ± 2	0.88 ± 0.04	0.83 ± 0.04
							Std. & D_0>65	98	6%	2	(2%)	63 ± 2	60 ± 2	31 ± 3	27 ± 2	0.91 ± 0.05	0.87 ± 0.04
							s_{Tn}<20% & D_0>65	150	9%	4	(3%)	61 ± 2	59 ± 2	33 ± 2	30 ± 2	0.88 ± 0.04	0.85 ± 0.04
							Std. & F_R>4	18	1%	1	(5%)	68 ± 7	65 ± 5	31 ± 6	27 ± 5	0.99 ± 0.10	0.93 ± 0.08
							s_{Tn}<20% & F_R>4	23	1%	3	(12%)	70 ± 5	67 ± 4	28 ± 5	24 ± 4	1.01 ± 0.08	0.97 ± 0.07
							Std. & D_0>65 & F_R>4	15	1%	1	(6%)	72 ± 8	68 ± 6	22 ± 4	16 ± 3	1.04 ± 0.12	0.98 ± 0.09
							s_{Tn}<20% & D_0>65 & F_R>4	20	1%	1	(5%)	73 ± 6	70 ± 5	27 ± 5	23 ± 4	1.06 ± 0.09	1.01 ± 0.08

[a] number of accepted aliquots.

[b] relative standard deviation

[c] number of measured aliquots.

[d] Std. is the standard rejection criteria given in section 3.3. s_{Tn}<20% includes grains with an uncertainty on the first (natural) test dose response known to better than 20%. D_0>x is the additional/alternative rejection criterion that the D_0 value of individual DRCs must be larger than x (see section 8.4). F_R>4 is the additional/alternative rejection criterion that the fast ratio, F_R, must be larger than 4 (see section 8.4).

[e] number of aliquots appearing to be in saturation.

[f] Relative over-dispersion calculated using CAM and uncertainties based on counting statistics and curve fitting errors (c.s.) and an additional uncertainty of 15% (c.s. +15%).

[g] Measured/Expected dose is the ratio between the SG and MG average doses.

Table 6. Summary of FMM results.

		Standard rejection criteria				Standard rejection criteria & D_0 criterion	
		σ_{add} = 15%		σ_{add} opt.		σ_{add} = 15%	
Sample	k	D_e (Gy)	p	D_e (Gy)	p	D_e (Gy)	p
092201	1	4 ± 2	0.01 ± 0.01	4 ± 2	0.01 ± 0.01	-	-
	2	**71 ± 2**	**0.52 ± 0.05**	72 ± 5	0.49 ± 0.10	79 ± 3	0.46 ± 0.07
	3	131 ± 4	0.47 ± 0.05	**124 ± 8**	**0.50 ± 0.10**	**138 ± 6**	**0.54 ± 0.07**
092202	1	10 ± 3	0.01 ± 0.01	10 ± 4	0.01 ± 0.01	-	-
	2	51 ± 6	0.09 ± 0.04	76 ± 8	0.43 ± 0.20	53 ± 7	0.07 ± 0.04
	3	**89 ± 4**	**0.58 ± 0.08**	**120 ± 11**	**0.57 ± 0.20**	**92 ± 5**	**0.58 ± 0.10**
	4	142 ± 8	0.33 ± 0.08	-	-	143 ± 10	0.35 ± 0.10
092203	1	30 ± 3	0.12 ± 0.03	30 ± 3	0.11 ± 0.03	41 ± 5	0.13 ± 0.08
	2	**64 ± 2**	**0.75 ± 0.05**	**64 ± 2**	**0.77 ± 0.05**	**65 ± 3**	**0.74 ± 0.08**
	3	117 ± 8	0.13 ± 0.04	117 ± 9	0.12 ± 0.04	118 ± 9	0.13 ± 0.05
092204	1	34 ± 3	0.18 ± 0.06	26 ± 5	0.06 ± 0.04	38 ± 5	0.15 ± 0.07
	2	**59 ± 2**	**0.72 ± 0.06**	**57 ± 2**	**0.86 ± 0.05**	**62 ± 3**	**0.77 ± 0.07**
	3	107 ± 8	0.10 ± 0.04	110 ± 16	0.08 ± 0.05	117 ± 9	0.07 ± 0.05

The standard rejection criteria are given in section 3.3 and the D_0 criterion is given in section 8.

k: component number

p: proportion of grains assigned to each component

σ_{add} = 15%: Individual uncertainties are based on photon counting statistics, curve fitting errors and an additional uncertainty of 15%

σ_{add} opt.: The additional uncertainty has been determined by optimising the BIC and llik scores. σ_{add} is 25, 22.5, 17.5 and 22% for samples 092201, -02, -03 and -04, respectively

Supplementary Data

Click here to download Supplementary Data: SI QUAGEO-D-13-00083R2.docx